International Journal of Computer Assisted Radiology and Surgery (2019) 14:545–561
https://doi.org/10.1007/s11548-018-1899-z

ORIGINAL ARTICLE

Segmentation of the proximal femur in radial MR scans using a random forest classifier and deformable model registration

Dimitrios Damopoulos[1] · Till Dominic Lerch[2] · Florian Schmaranzer[2] · Moritz Tannast[2] · Christophe Chênes[3] · Guoyan Zheng[1] · Jérôme Schmid[3]

Received: 19 February 2018 / Accepted: 10 December 2018 / Published online: 2 January 2019

Abstract

Background Radial 2D MRI scans of the hip are routinely used for the diagnosis of the cam type of femoroacetabular impingement (FAI) and of avascular necrosis (AVN) of the femoral head, both considered causes of hip joint osteoarthritis in young and active patients. A method for automated and accurate segmentation of the proximal femur from radial MRI scans could be very useful in both clinical routine and biomechanical studies. However, to our knowledge, no such method has been published before.

Purpose The aims of this study are the development of a system for the segmentation of the proximal femur from radial MRI scans and the reconstruction of its 3D model that can be used for diagnosis and planning of hip-preserving surgery.

Methods The proposed system relies on: (a) a random forest classifier and (b) the registration of a 3D template mesh of the femur to the radial slices based on a physically based deformable model. The input to the system are the radial slices and the manually specified positions of three landmarks. Our dataset consists of the radial MRI scans of 25 patients symptomatic of FAI or AVN and accompanying manual segmentation of the femur, treated as the ground truth.

Results The achieved segmentation of the proximal femur has an average Dice similarity coefficient (DSC) of $96.37 \pm 1.55\%$, an average symmetric mean absolute distance (SMAD) of 0.94 ± 0.39 mm and an average Hausdorff distance of 2.37 ± 1.14 mm. In the femoral head subregion, the average SMAD is 0.64 ± 0.18 mm and the average Hausdorff distance is 1.41 ± 0.56 mm.

Conclusions We validated a semiautomated method for the segmentation of the proximal femur from radial MR scans. A 3D model of the proximal femur is also reconstructed, which can be used for the planning of hip-preserving surgery.

Keywords Radial imaging of the hip · Proximal femur · 3D reconstruction · Segmentation · Random forest · Deformable model

Introduction

Femoroacetabular impingement (FAI) and avascular necrosis of the femoral head (AVN) are known causes of osteoarthritis of the hip joint in young and active patients [1]. FAI has been described in [2] as a biomechanical entity that originates from anatomical abnormalities of the proximal femur and/or the acetabulum which manifests itself with decreased range of motion and pain [3–7]. FAI is the main cause of early onset osteoarthritis in non-dysplastic hips [4]. In the cam type of FAI, anatomical abnormalities are observed only on the femoral head without any pelvis involvement. On the other hand, in AVN the blood flow to the femoral head is interrupted, which can progressively lead to the collapse of the hip. A lot of joint-preserving forms of treatment have been

Guoyan Zheng and Jérôme Schmid share the senior authorship

✉ Dimitrios Damopoulos
dimitrios.damopoulos@istb.unibe.ch

✉ Guoyan Zheng
guoyan.zheng@istb.unibe.ch

[1] Institute for Surgical Technology and Biomechanics, University of Bern, Stauffacherstrasse 78, 3014 Bern, Switzerland

[2] Department of Orthopaedic Surgery and Traumatology, Inselspital, University of Bern, Freiburgstrasse, 3010 Bern, Switzerland

[3] School of Health Sciences - Geneva, HES-SO University of Applied Sciences and Arts Western Switzerland, Avenue de Champel 47, 1206 Geneva, Switzerland

developed in an attempt to slow or reverse its progression, as it usually affects young patients [1, 8]. MRI has been recognized as an important assisting tool for the diagnosis and the assessment of FAI and AVN as, in addition to the non-ionizing nature of its radiation, MRI can capture the vascular status of the femoral head [9–12]. Moreover, as MR scanners typically have the capability to directly scan planes of arbitrary orientation, it is possible to acquire images perpendicular to the curvature of the acetabulum, as visualized in Fig. 1. Such a scanning protocol is often referred to as *radial imaging of the hip*. The appeal of using radial scans over 3D MRI for image-assisted diagnosis is the reduced scanning time, as a typical radial scan of the hip consists of much fewer slices. Indeed, acquisition of 3D MR images is typically not part of the clinical routine, as MR scanning time tends to be a resource high in demand. Additionally, radial imaging, using a gadolinium-enhanced scanning protocol [13–16], is considered the gold standard for diagnosis of FAI and AVN [1, 17, 18].

In addition to the value of MRI in diagnosis, MR-based 3D models of the femur have recently been shown to be as accurate as CT-based ones [19, 20]. Such 3D models form an essential part of the planning of hip-preserving surgeries, and they are also valuable in conducting motion analysis studies. In particular, surgical correction of cam-FAI or AVN is challenging and requires exact preoperative planning [5–8]. The reconstruction of such a 3D model requires the segmentation of the femoral area in every slice, a task which, if performed manually, is very tedious. In the case of MR radial scans, the effort required is much lower. Still, the segmentation result is susceptible to inter-observer variability. Moreover, radial scans can suffer from a crosstalk artifact at the intersection of all the images and it can be unclear how to segment the proximal femur at the affected locations.

For these reasons, an automated and accurate method for the segmentation of the proximal femur in radial scans of the hip has the potential to increase the value of this type of sequence in both clinical routine and biomechanical studies. In this work, we present a novel method for this task. The proposed method registers a 3D template mesh of the femur to the radial slices based on a physically based deformable model. The registration process utilizes the pixel-wise predictions of a classifier. An evaluation study is conducted on a dataset of 25 radial scans of symptomatic patients.

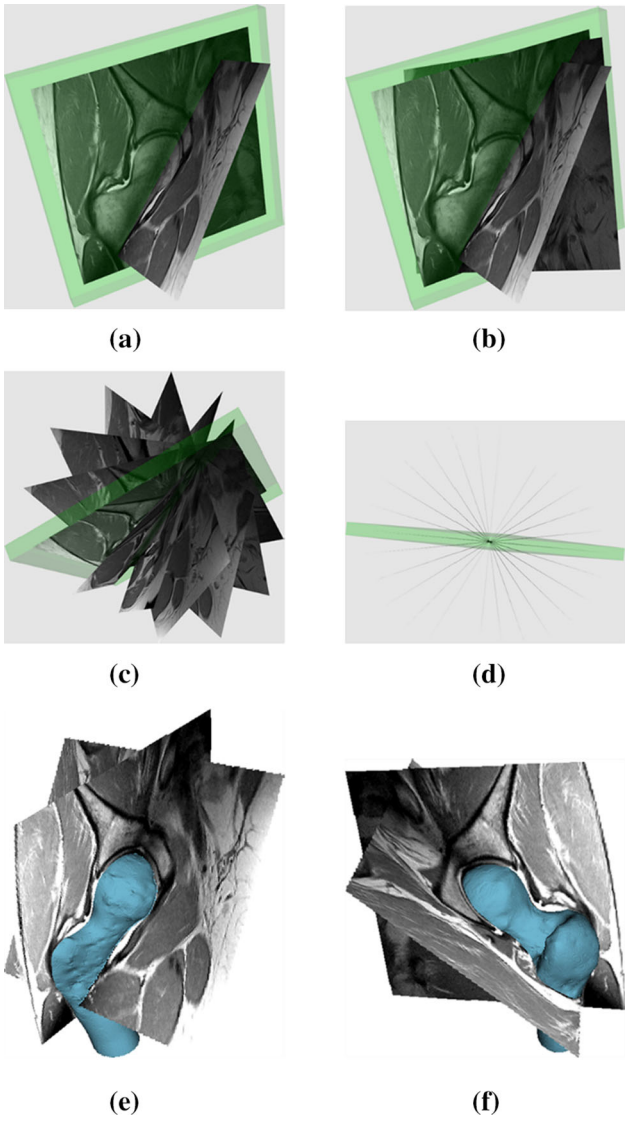

Fig. 1 Visualization of the 3D geometry of a radial scan along with the registered 3D model, shown from different 3D viewpoints. In **a–d,** the same image is highlighted with a green overlay. For clarity of visualization, only two, three and seven of the total 14 images of the radial scan are shown in **a–c,** respectively. In **d,** it is can be observed that all the images share a common axis and that their angular displacements around their common axis are uniformly distributed around the circle. Finally, in **e** and **f**, the two images of **a** are shown along with the registered 3D model, as generated by the proposed method. Subfigures **e** and **f** were created using the 3D Slicer software [41]

Related work

There are a few published methods on the problem of the automatic segmentation of the proximal femur from 3D MR images. In [21], the authors applied both multi-atlases and active shape models (ASMs) for the segmentation of the proximal femur and the hip bone. In [22–25], different methods based on deformable models were proposed: In [22], a robust multi-resolution statistical shape model (SSM) method for the segmentation of bones in small field of view (FOV) was presented and results for the case of the femur and of the hip bone were reported. In [23], a general segmentation method for both muscles and bones was presented and an evaluation for femur segmentation was conducted. In [24], user-specified landmarks were also utilized for the segmentation of the proximal femur. In [25], weighted shape prior was introduced to the deformable model. Whereas all

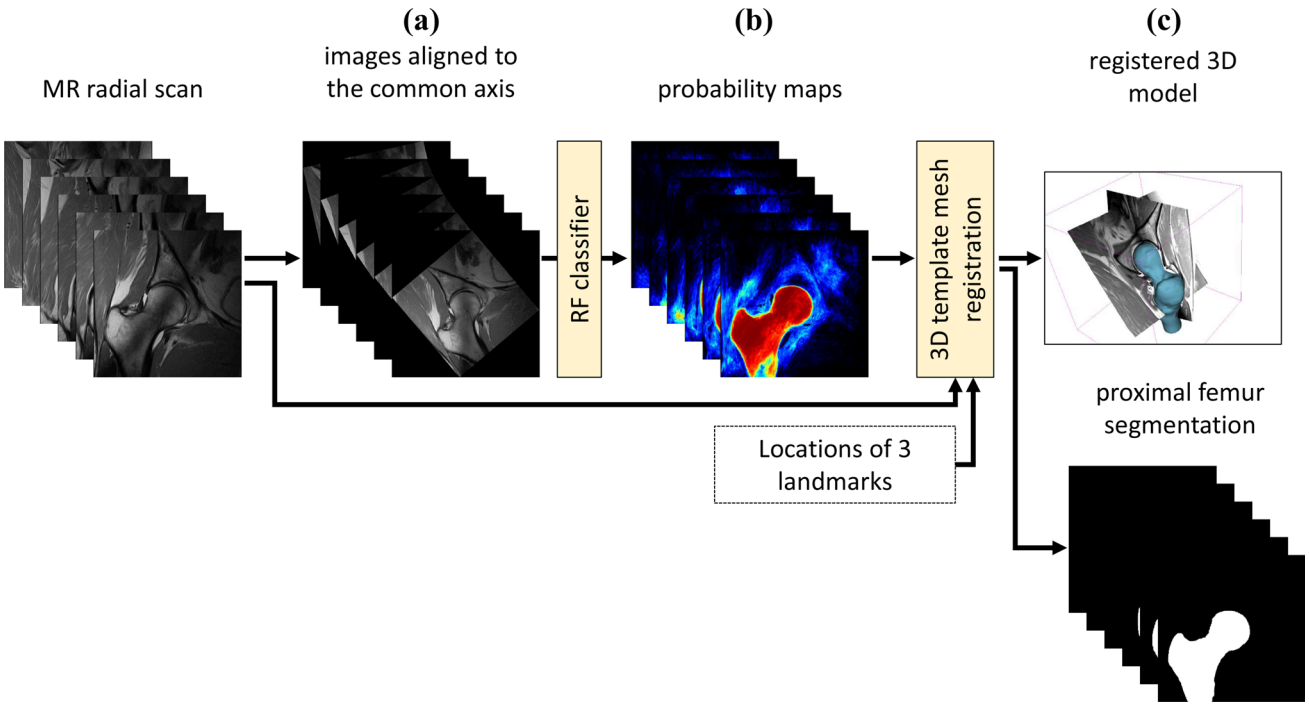

Fig. 2 Summary of the pipeline. **a** The images of a radial scan are preprocessed, and they are aligned according the common axis of the scan; **b** the random forest classifier makes a prediction for every pixel of the images on whether it corresponds to femur. The generated probability maps are rotated back to the original orientations; **c** a 3D template mesh of the femur is registered to the radial images using the generated probability maps to drive the registration process and the positions of the three landmarks to initialize it

the previous methods are model-based, in [26] the authors presented a purely classification-based approach, by proposing a 3D deep learning network architecture based on u-net for this classification task.

All the methods mentioned above assume that 3D MR data are available. To the best of our knowledge, no method for the segmentation of the proximal femur has been proposed that relies solely on radial scans. Turning our attention to different organs, we can find the literature on the segmentation of heart from similar radial images, for example, [27, 28]. As summarized in [27], these methods tend to have an interpolation part and a segmentation part. Sometimes, the slices are segmented independently from one another and the 3D surface is interpolated by the 2D contours. Alternatively, often a 3D volume is firstly interpolated from the 2D slices, and then, some segmentation method, usually model based, is applied on the 3D volume.

The geometric feature that we are using in the classification stage is similar to those of the auto-context framework [29], especially to the distance-based features introduced in [30]. The main difference with these works is that they employ at least one prior regression or classification layer in order to locate certain landmarks or objects. In the present work, we attempt to directly take advantage of prior knowledge concerning the orientation of the images without a prior localization step.

Method

The input to the pipeline is a set of radial MR images of the hip. Their FOV includes the hip joint and the upper extremity of the femur. In the following sections, we will be referring to such an MR sequence as *a radial scan* or simply *a scan*. For the individual planes of a radial scan, we will use the term *images*.

Unlike typical MR scans, the images of a radial scan are not parallel to each other. Instead, they lay on oblique planes, acquired in a radial (rotating) fashion along the axis of the femoral neck. Their acquisition protocol is described in detail in [13], which specifies 14 images per scan. Following this protocol, the images are gadolinium-enhanced. The geometric arrangement is visualized in Fig. 1, which presents an example radial scan from different 3D viewpoints.

The proposed pipeline consists of: (a) a preprocessing step, where the input images are aligned; (b) a classification step with a random forest classifier that estimates the probability of every pixel to belong to femur; (c) a deformable model registration stage which fits a 3D template mesh to the input images and the generated probability maps. A summarizing diagram is presented in Fig. 2. Each component is described in the following sections.

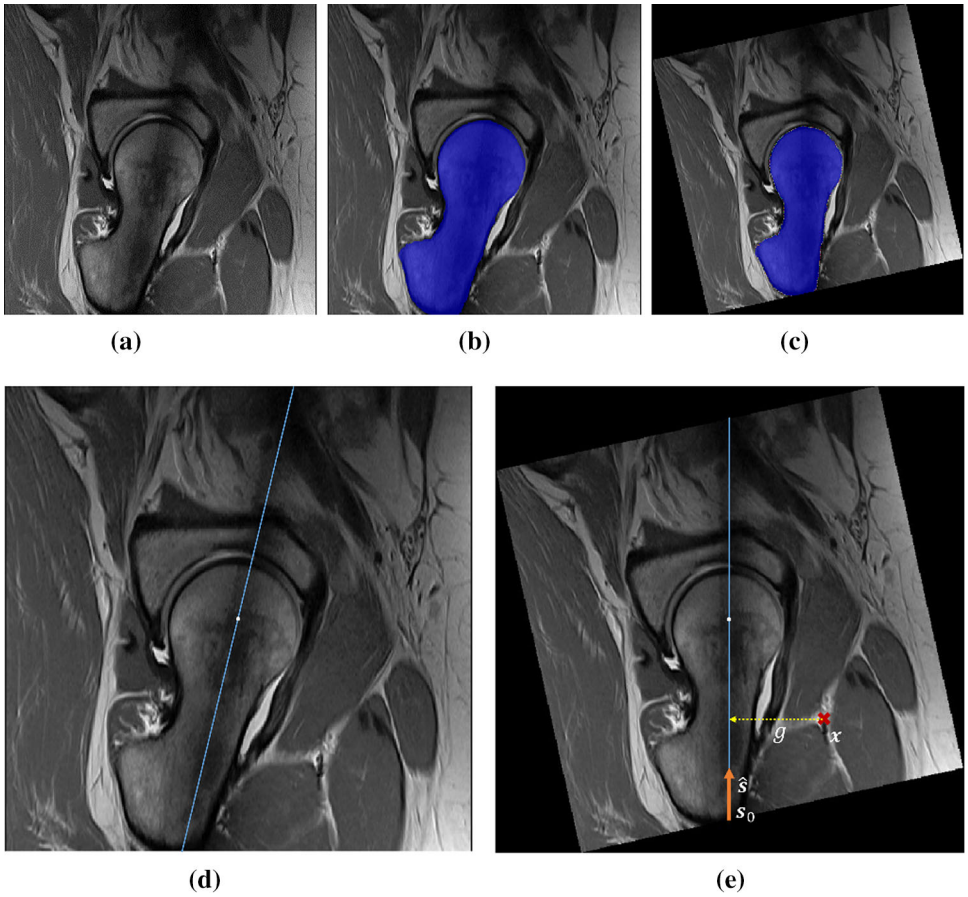

Fig. 3 Alignment of the images with respect to the common axis and visualization of the geometric feature. From **a** to **c**: An example input image; the same image, with the ground truth segmentation of the femur highlighted with a blue overlay; the appearance of the image, when it is aligned with respect to the common axis (the image has been rescaled on this figure to make the figure clearer). The crosstalk artifact, occurring on the common axis of all the images and typical for MR radial scans, can be easily observed. **d** The image of subfigure (**a**), with the common axis pinpointed with a blue line. **e** The aligned image of subfigure (**c**), with the common axis pinpointed with a blue line. The unit vector $\hat{s}$ is the direction vector of the common axis. Since the image is aligned, $\hat{s}$ is parallel to the vertical axis. The length of the yellow dotted line is the value of the geometric feature g, computed at the point $\boldsymbol{x}$

Preprocessing

At the preprocessing stage, two operations are carried out: (a) The intensities of the images are normalized by histogram matching to a reference image, and (b) the images are rotated so that the common axis is vertical on every aligned image.

The intensity normalization is performed because the range of the intensities of an image can be different for different radial scans, which might affect the performance of the intensity-based features that are used by the random forest classifier. The matching is performed using a standard algorithm for histogram matching for MR images [31].

In order to reduce the variability of the appearance of the femur in the images and to assist the classification stage that follows, the images are rotated so that the common axis is always vertical. This is accomplished in the following manner: Firstly, the common axis is retrieved by computing the intersection of any two images in the scan. Then, the orientation of the common axis with respect to every image is calculated. The desired rotation angle is the opposite of the angle of the computed orientation. The result of this alignment procedure is shown in Fig. 3.

The locations of three anatomical landmarks are also specified manually by the user, to be used for the initialization of the deformable model registration procedure of the last stage.

Random forest classifier

The automated segmentation is based on a binary, pixel-wise classifier which makes a probabilistic estimate for every pixel of an aligned image on whether it belongs to the femur area or not. These two classes of pixels will be referred to as *foreground* and *background* in the following sections.

A random forest is utilized for the binary classification. In addition to standard, intensity-based features, a geometric feature is used that attempts to introduce spatial context information to the classification process. For the description of these features that follows, it is assumed that an image $I{:}\Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ is given, with the domain Ω corresponding to the pixel coordinates of the image. Since the classifier treats the images of a scan independently from each other, we ignore the fact that the pixels correspond in fact to 3D points in the world coordinate system and, for this subsection, we treat the images as 2D objects.

Intensity-based features

A standard type of intensity-based features is utilized, the same way as in [32, 33, 44–46, 49]. These features are based on the mean intensity value over displayed boxes. In more detail, if $\boldsymbol{x} \in \Omega$ is a reference point, the following features are considered:

$$f(\boldsymbol{x}; \boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{o}_1, \boldsymbol{o}_2, s) = \frac{\sum_{y \in B_1} I(\boldsymbol{x} + \boldsymbol{o}_1 + \boldsymbol{y})}{|B_1|} - s \frac{\sum_{y \in B_2} I(\boldsymbol{x} + \boldsymbol{o}_2 + y)}{|B_2|}$$

where $s \in \{0, 1\}$, $\boldsymbol{o}_1, \boldsymbol{o}_2$ are 2D offsets, $\boldsymbol{B}_1, \boldsymbol{B}_2$ are 2D rectangular boxes and $|B_1|, |B_2|$ are their areas.

As in the previously mentioned publications, a pool of these features is sampled randomly from a predefined range of values for $\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{o}_1, \boldsymbol{o}_2$ at the beginning of the random forest training. The interested reader is referred to [46] for a detailed description of this family of features and how they are integrated in the random forest classification framework.

A geometric feature

In order to improve the performance, we take advantage of prior knowledge concerning the geometric arrangement of the images in a scan. As described earlier, the images of a radial scan of the proximal femur intersect on one axis in the 3D world coordinates space which passes close to the femoral head center and the femoral neck center. This prior knowledge is incorporating in the classifier through a feature which is simply the distance of the reference point to the common axis. This can be expressed as:

$$g(\boldsymbol{x}; \boldsymbol{s}) = \left\| (\boldsymbol{x} - \boldsymbol{s}_0) - \left((\boldsymbol{x} - \boldsymbol{s}_0)^{\boldsymbol{T}} \cdot \hat{\boldsymbol{s}} \right) \hat{\boldsymbol{s}} \right\|_2$$

where $\boldsymbol{s}_0$ is a point of the common axis and $\hat{\boldsymbol{s}}$ is the direction vector of the common axis. This is more intuitively illustrated in Fig. 3e. As it can be seen on this figure, this distance is calculated on the aligned image, so the unit vector $\hat{\boldsymbol{s}}$ is simply equal to $[0, 1]^T$ and the computation above simplifies to just the subtraction of the y coordinate of $\boldsymbol{s}$ from the y coordinate of $\boldsymbol{x}$.

Deformable model fitting

In the last stage, a 3D template mesh of the femur is registered to the radial images based on a physically based deformable model. The implemented deformable model framework is partially based on our previous work of physically-based simplex meshes [22, 34]. In this framework, mesh vertices are considered as lumped mass particles whose motion is driven by forces and which follows Newtonian dynamics. By carefully crafting the forces, the deformable model will reach an equilibrium position corresponding to the structure to segment.

The state of a particle i at time t is described by its position $P_i(t)$ and velocity $P_i^{'}(t)$. Following Newton's law of motion, the particle acceleration $P_i^{''}(t)$ is related to the inverse of the particle mass, expressed by its inertia matrix L_i, and the sum of forces F_i exerted: $P_i^{''}(t) = L_i F_i(t)$. The particle motion is thus described by a differential equation system which can be linearized and solved according to an integration scheme. We chose the implicit Euler scheme as described in [35]:

$$\begin{aligned} P(t + \mathrm{d}t) &= P(t) + P'(t + \mathrm{d}t)\mathrm{d}t \\ P'(t + \mathrm{d}t) &= P'(t) + L^T H^{-1} Y \\ H &= L^T - L \frac{\partial F}{\partial P'} L^T \mathrm{d}t - L \frac{\partial F}{\partial P} L^T \mathrm{d}t^2 \\ Y &= L F(t) \mathrm{d}t + L \frac{\partial F}{\partial P} P'(t) \mathrm{d}t^2 \end{aligned}$$

This scheme remains very stable when using large integration time steps and can be efficiently implemented by using the conjugate gradient algorithm and by exploiting the sparsity of the system matrix H. A simulation step consists of: (a) computing forces and their derivatives and (b) updating the next particle state by solving the differential equation system. For a vertex at position P_i, a force f_i is commonly modeled as the force of a Hookean spring which will attract the vertex toward a *target* vertex position R_i:

$$f_i = \propto_i (R_i - P_i)$$

where $\propto_i$ is the weighting force coefficient.

A key difference with respect to our previous framework [22, 34] is that we do not rely anymore on simplex meshes. We use instead standard triangular meshes due to the maturity and availability of many geometrical modeling techniques for this category of meshes. As a result, we devised new approaches in the framework, as detailed in the following paragraphs.

Internal forces

Internal forces regulate mesh deformation by enforcing *smoothness* and shape similarity. Shape similarity is enforced by *local* and *global shape* forces. Smoothness and local shape forces focus on a local geometric description of a vertex with respects to its neighbors, while strong shape forces are based on statistical shape models [36] and affect the mesh globally.

Smoothing and local shape forces

Contrary to triangular meshes, simplex meshes are characterized by constant vertex connectivity, simplifying the

computation of various local geometrical properties that were used in our previous framework.

Instead of relying on the dual transformation between simplex and triangular meshes, we choose in the present work to model the vertex local geometry with the Mean Value Encoding (MVE) of [37]. In MVE, for each vertex at position P_i a local plane $\Pi_i(n_i, d_i)$ is built based on the m_i neighbor vertex positions Q_j, $j \in N_i$. The plane normal n_i is computed as an area-weighted sum of the neighbor vertex normals, where the area a_i of a vertex i is computed from the areas A_{ik} of the triangles sharing the vertex: $a_i = \sum A_{ik}/3$. The plane position d_i is expressed as an average distance from the origin:

$$d_i = -\frac{1}{m_i} \sum_{j \in N_i} n_i \cdot Q_j$$

Each vertex position is expressed in terms of tangential and normal components with respect to the local plane, by a series of MVE parameters. These parameters provide an efficient encoding of local geometry at each vertex, robust to degenerate situations such as nearly collinear neighbor vertices.

MVE parameters can be precomputed for the mesh at $t = 0$, providing a "snapshot" of the mesh shape. Then, during mesh deformation, for each vertex position P_i a decoding process computes the position R_i that would be expected based on the MVE encoding. The local shape force uses this position R_i as the target position to enforce local shape geometry.

The MVE local geometry can also be used to design a smoothing force preventing excessive local deformations such as spokes. The target point R_i for vertex i of the smoothing force is:

$$R_i = \sum_{j \in N_i} a_j (Q_j + h_j) / \sum_{j \in N_i} a_j$$

where h_j denotes the *height* of the vertex j in its local geometry, i.e., the normal component of position P_j with respect to its local plane Π_j. This smoothing is similar to Laplacian smoothing which uses the vertex barycenter, but it additionally reduces the shrinking effect of the Laplacian smoothing by including the weighted height.

Global shape force

In [22], we presented the use of statistical shape models (SSM), expressed as point distribution models [36], to create the global shape force. The idea is to compute a shape S based on the SSM which is the closest in least squares sense to the current mesh M. This is based on an iterative procedure described in [36] which estimates the best alignment transform and shape parameters of the SSM yielding the closest shape S. The vertices of S are eventually used as the target vertices R_i for the global shape forces (Fig. 4).

External forces

External image forces use image cues to guide the deformation of the mesh and are specifically designed based on the modality and nature of the images. In this work, we use both pixel intensities of the MR images and the values of the probability maps of the random forest classifier. Without any loss of generality, we will refer to any of them as an image with intensities I.

Given the 3D plane of a radial image, we select the mesh vertices P_i^* whose projection on the plane L_i^* is within the image bounds and whose distance to the plane is below a very small threshold l (Fig. 5a). Then, the normals of the vertices P_i^* are projected on the plane as 2D vectors n_i^*. In the plane, we sample $2W + 1$ values every s mm along the normal direction n_i^* in an interval centered on L_i^*. The sampling position u where the intensity gradient $\nabla I(u)$ has *(i)* the greatest magnitude but also *(ii)* the closest direction to the projected normal n_i^* is chosen for the force target point R_i (Fig. 5b):

$$R_i = \text{argmax } \varepsilon * \nabla I(u) \cdot n_i^*$$

where ε equals +1 or -1 depending on whether the normals need inversion (in the case of probability maps, outward mesh normals and image gradients will have opposite directions). This force will only affect a subset of mesh vertices P_i^* due to the small FOV of the radial images. The remaining vertices are ignored during the alignment procedure and the least square minimization of the global shape force computation [22].

Regularization

Internal forces play an important role in preventing incorrect mesh deformations due to image forces affected by image noise or neighbor anatomical structures. However, additional regularization strategies are generally required to bring robustness and tackle possible numerical issues.

First, we adopt the same multi-resolution strategy depicted in [22], where several resolutions of the mesh are successively used during the mesh deformation. We also use a coarse-to-fine SSM scheme, by progressively increasing the SSM "locality" [34] during the process. The locality is related to the alignment type of the SSM, and it is an intuitive notion: A rigid SSM captures global shape changes while an affine or a similarity SSM will better express local variations.

Second, we used a damping force $f_i = - \propto_i P_i^{'}$ to prevent possible instabilities of the integration scheme. This may occur during the approximation of the force derivatives or

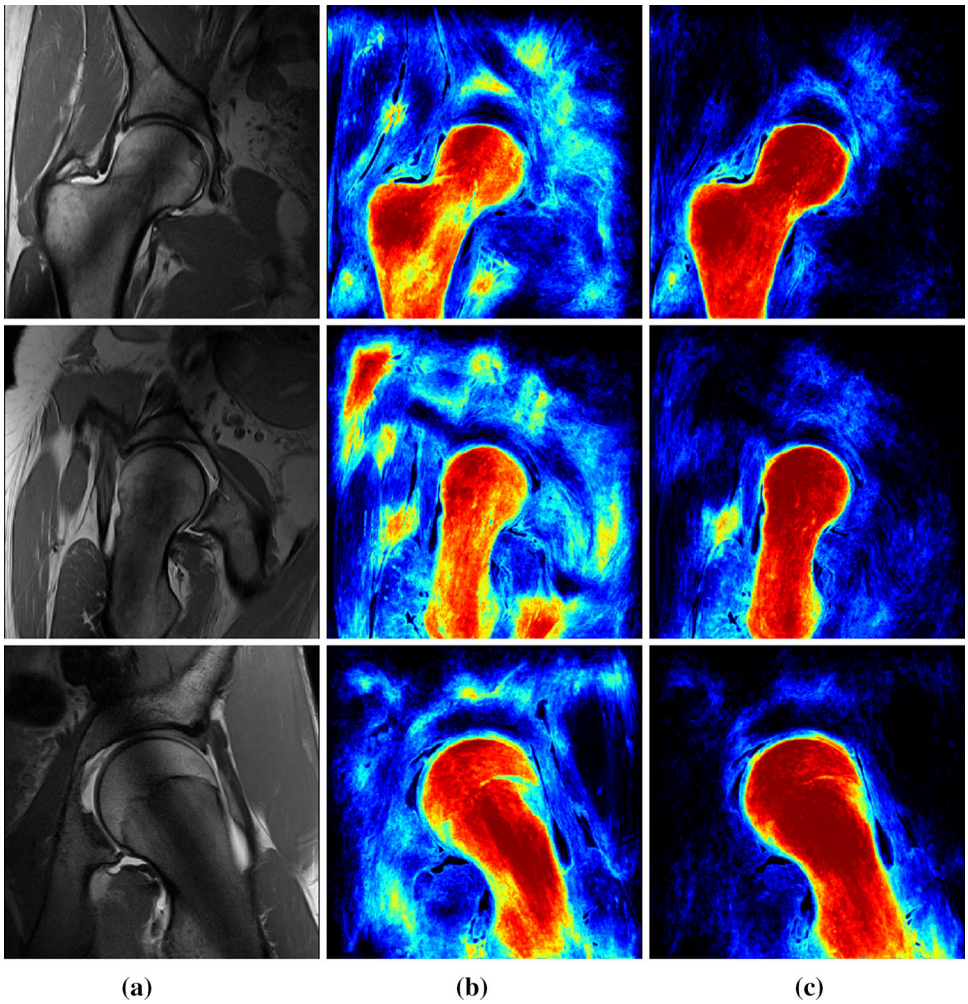

Fig. 4 Probability maps with and without using the geometric feature. **a** Three testing images; **b** the probability maps generated by a classifier that used only the intensity-based features; **c** the probability maps generated by a classifier that used both the intensity-based and the geometric features. The probability maps were visualized using the ITK-Snap software [40]

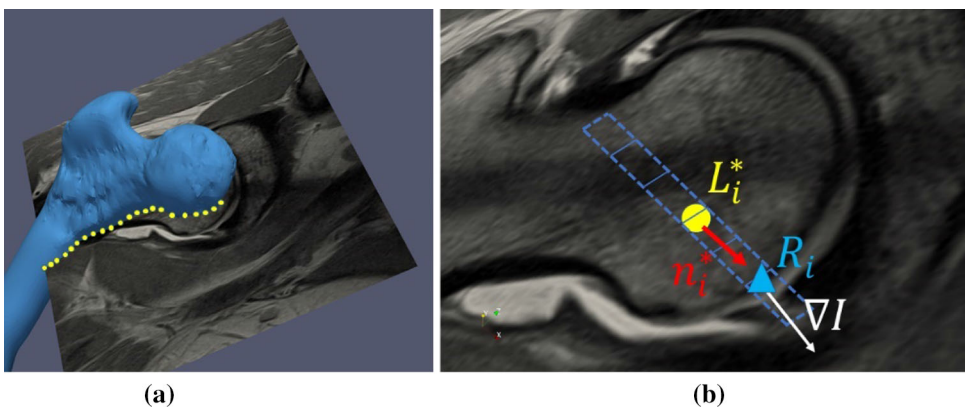

Fig. 5 Search of target point R_i for the image force. **a** Mesh vertices close enough to the image plane and whose projection L_i^* is within image bounds are selected (yellow ●). **b** For each vertex projection L_i^* (●), values are sampled along the projected normal direction n_i^* at regular steps s. The sampled position at which the image gradient ∇I is the largest and the best aligned with n_i^* is chosen as the target point R_i (▲)

when particle interactions are ignored. Approximate derivative may be chosen to preserve the system symmetry required by the conjugate gradient method or to simplify computations. For instance, the global shape force computation for a vertex i will actually depend on *all* particle positions at the previous time step. However, we intentionally (wrongly) assume that the resulting R_i will not depend on the particle positions. In practice, neither instabilities nor odd behaviors were noticed with the chosen approximations.

Experimental design and results

Dataset

Our dataset consists of 25 MR gadolinium-enhanced radial scans of 25 patients symptomatic of FAI or AVN. They were acquired in the period 2010–2016 in the Sonnenhofspital, Bern, Switzerland, following a radial sequencing acquisition protocol of the hip joint [13]. The age of the patients is 16–47 years with a mean age of 29.0 years, 11 are male, and 14 are female. In 12 patients it is the left hip that is scanned and in 13 cases it is the right one. The 3D geometry of an example radial scan is visualized in Fig. 1. The intra-slice spacing is in 0.28–0.29 mm range, the size of the images is either 448×448 (19 cases) or 512×512 (6 cases), and there are 14 images in every scan. Radial scans of this type are routinely acquired in the aforementioned hospital. A reference, manual segmentation of every image of the radial scans is also provided.

Study design

A fivefold cross-validation study is performed on the dataset of the 25 radial scans with every fold consisting of five scans. Therefore, each scan is utilized exactly once as part of a test set. The reference image of the intensity normalization preprocessing step is set to be the image of the training set with the median mean intensity. For the initialization of the deformable model registration stage, the following three landmarks were used: (a) the femoral head center, (b) the femoral neck center and (c) the tip of the minor trochanter. These were specified manually for each case using the "Fiducials" module of the Slicer 3D open-source software [41].

Hyper-parameters

Random forest

Given the relatively small size of our dataset, the most preferable way to tune the hyper-parameters of the random forest would be independently for every cross-validation iteration through a nested cross-validation scheme. However, due to the long training times (more than half a day for one iteration), we opted to set the hyper-parameters in advance to fixed values, based on related literature and prior knowledge on the task at hand. Specifically, when random forests are utilized as classifiers for image segmentation problems (as in [42–46]), the maximum depth of the trees is usually in the $T \in [20, 40]$ range and their number in the $N \in [20, 120]$ range. These parameters affect the generability of the trained classifier: For a fixed depth, more trees reduce the variance of the model (but lead to higher training and testing times) and, for a fixed number of trees, the depth influences the bias–variance trade-off. These effects of T, N can be explained theoretically [50], and they have been observed in practice [46–48]. We conservatively set the maximum depth to $T = 20$ to prevent overfitting and the number of trees to $N = 50$. The box size of the intensity-based features is between 0.3 mm, which is slightly larger than the pixel spacing in the dataset, and up to 7.5 mm, as we expect that no area in the foreground larger than 7.5×7.5 mm^2 has uniform intensity. Following the practice of our prior work [49], we set the maximum range of the intensity features to double the maximum length of the boxes, i.e., to 15 mm.

A posteriori, we varied the tree depth as $T \in \{15, 20, 25, 30\}$, keeping the number of trees to $N = 50$. We found out that for T = 25 for the classifier-based method, the mean scan DSC increased by 0.07% and the mean SMAD decreased by 0.02%, suggesting that there is some room for better tuning of our method, probably though only marginal. Since we did not follow a nested cross-validation strategy, we are reporting here the results obtained with the preset parameter values.

Deformable model

The hyper-parameters of the deformable model registration stage were chosen based on an empirical analysis performed on three randomly chosen cases. The procedure was based on a semiautomatic interactive segmentation, during which a skilled operator could increase or decrease the influence of forces as well as increasing the resolution of the mesh when needed. Some initial parameters of the segmentation, such as weight forces or image force coverage, were derived from our previous experience in MRI segmentation of the hip [22]. From this previous work, we could already identify some best strategies to apply, such as the choice of the SSM alignment with respect to the mesh resolution. All operator actions were recorded in a script which could be reapplied later on in an automatic fashion on the test cases. If some high errors were detected, the operator would rerun the interactive segmentation and adapt her/his choices until a satisfactory segmentation was observed for the three cases. As a measure against overfitting with respect to these cases, the script was cleaned out to keep the parameters as constant as possible, since the operator varied some of them unnecessarily over time. The parameters whose variation was found to have a significant impact on the results were the global shape force weight and image force coverage; only these two were allowed to evolve over time.

Using the strategy described above, the deformable model used 4 mesh resolutions, ranging from 700- to 50-K points. The mass of a vertex was set as the total surface of the mesh divided by the number of vertices. Mesh resolution 3 (the coarsest) used a rigidly aligned SSM, while resolutions 2 and 1 used an SSM with affine alignment. The highest resolution

0 did not use any global shape force. The weighting force coefficients $\propto_i$ of the smoothing and local shape forces were, respectively, set to 0.2 and 0.3. The global shape force weight progressively decreased from 0.8 to 0.1 during the simulation as follows: For 51% of the total iterations count the weight was set to 0.8, then to 0.4 for 8%, 0.3 for 13%, 0.1 for 20% and 0 (the force being not used for the highest mesh resolution) during the remaining 8%. The statistical shape models were built from training femur shapes produced by segmenting a collection of 200 hip CT images. The image force weight was kept constant at 0.05 for the MR image and 0.45 for the probability map, but the image force coverage decreases over time from 9 ($W = 4$) samples to 5 ($W = 2$) samples spaced by $s = 1.2$ mm. The distance threshold l of the image force was set to 0.001 mm. Finally, damping weight was set to 0.5 and the time step was 1.0.

Evaluation metrics

The accompanying manual segmentation of the proximal femur is treated as the ground truth in our study. We compare the generated segmentation with this reference segmentation with respect to area overlap and border distance.

The area overlap is quantified with the Dice similarity coefficient (DSC), and two types of DSC measurements are performed: (a) The DSC is computed independently for every 2D image, and (b) all the images of a scan are treated as a single set and their DSC is computed with respect to the set of the image segmentations.

The two borders are compared using the symmetric mean absolute distance (SMAD) measure and the Hausdorff distance measure. The segmentation of every radial scan is treated as a single set; thus, all the images of a scan are taken into consideration. We also repeat these measurements specifically for the femoral head area by isolating the region from the hip joint to the femoral head center using a 3D mask.

The achieved scores on these metrics are presented in Table 1. In Fig. 6, the images of the radial scan with the median scan DSC value (96.65%) are listed, overlaid with the reference segmentation and the output of our method. Subfigures e and f of Fig. 1 illustrate an example registered 3D model of a testing case, along with two of its radial images.

When the three cases that were utilized for the configuration of the deformable model stage are omitted from the evaluation, the mean SMAD gets slightly better (0.62 mm instead of 0.64 mm), the scan DSC gets slightly worse (96.34% instead of 96.40%), and the mean values of the remaining 4 evaluation metrics stay the same, suggesting that no significant bias was introduced by the inclusion of these three cases in the evaluation.

Table 1 Performance of the proposed pipeline

	Mean	SD	Median	Min.	Max.
DSC over scans (%)	96.37	1.55	96.65	90.90	97.47
DCS over images (%)	96.40	1.45	96.68	89.86	97.40
SMAD (mm)	0.94	0.39	0.84	0.69	2.20
SMAD, only femoral head (mm)	0.64	0.18	0.54	0.40	1.06
Hausdorff (mm)	2.37	1.13	2.14	1.43	4.35
Hausdorff, only femoral head (mm)	1.41	0.56	1.28	0.81	3.83

The DSC over whole scans (each one consisting of 14 images) and the DSC over individual images are both presented
The symmetric mean absolute distance (SMAD) and the Hausdorff distance are computed for both the whole proximal femur and for the femoral head area only

Evaluation of specific components of the pipeline

The study was repeated for two variants of the proposed method, using the same five folds, in order to quantify the effect of specific components of the pipeline.

Firstly, we assess the effect of the deformable model registration stage on the segmentation performance by comparing the proposed pipeline with an alternative one that is based on the classifier without any registration of a 3D model. In order to make the comparison fair, we improve the result of the alternative method with standard post-processing operations: Firstly, a Conditional Random Field (CRF) [38] with a simple Potts model is applied, whose unary potentials are the probabilistic prediction of the random forest. The CRF inference is performed using the DGM C++ library [51]. Then, any holes on the resulting segmentation are filled and its largest connected component is the final output. The performance of this classifier-based method is summarized in Table 2.

Secondly, the importance of the geometric feature is assessed by measuring the segmentation performance when this feature is not utilized. The scores achieved are presented in Table 3.

Statistical tests were conducted for each of the two variants in order to determine whether their performance differs significantly from the default pipeline, in the sense that is described in this paragraph. In particular, for each of the six evaluation metrics, a *two-sided Wilcoxon signed-rank test* was performed, hereinafter abbreviated as WSR. Let $m_i, v_i, i \in \{1, \ldots, N\}$ be the measurements of two methods M, V, as obtained on a set of N samples. Let also D_M, D_V denote the distributions that m_i, v_i, respectively, are sampled from. WSR examines the differences $x_i = m_i - v_i$, whose

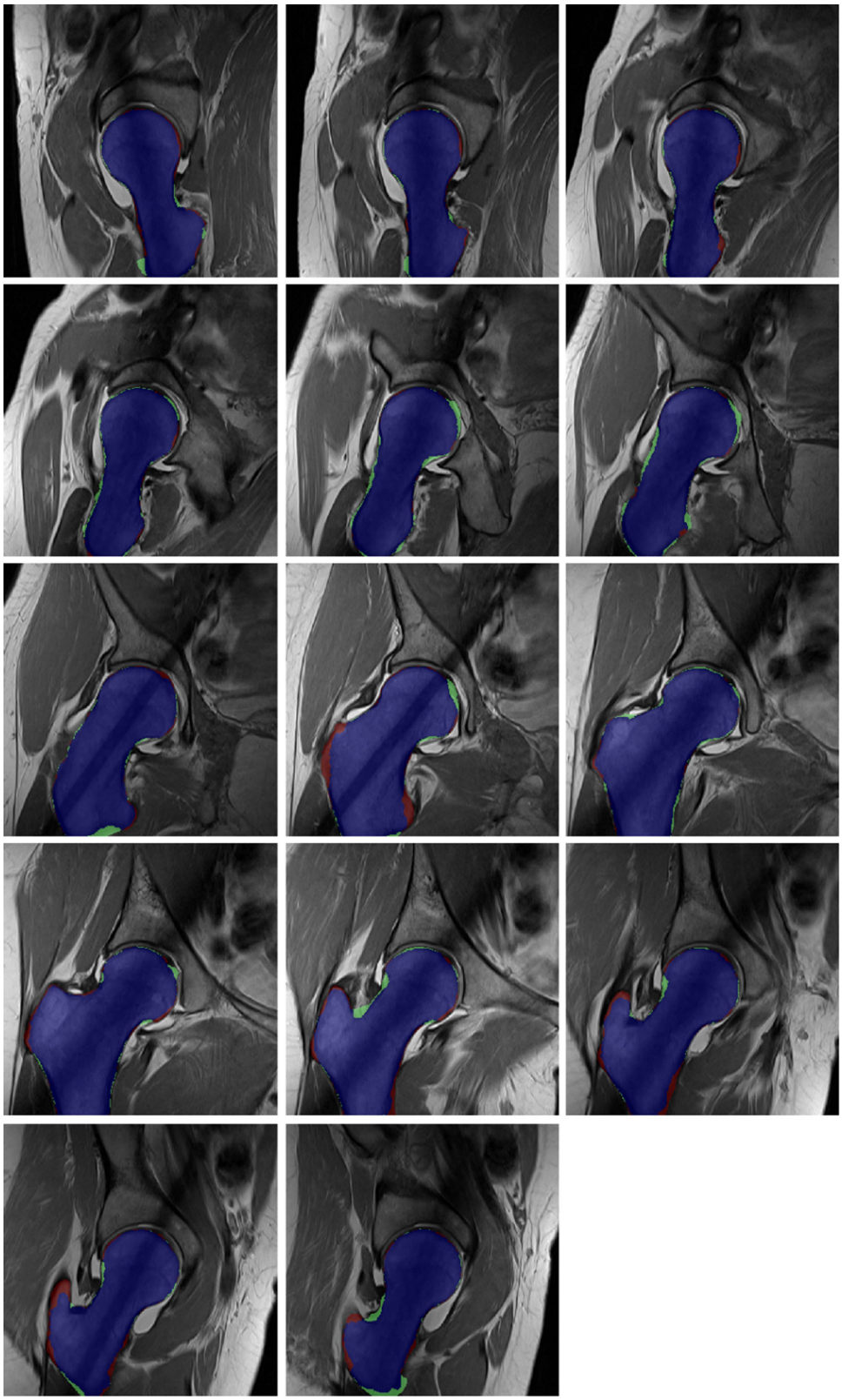

Fig. 6 Segmentation result for the 14 images of the case with the median DSC (96.65%) and illustration of its errors with respect to the manual, reference segmentation. Blue: true positive pixels; red: false negative pixels; green: false positive pixels

distribution is denoted with D_X. WSR makes the following assumptions: that the N samples are chosen randomly, that the values of the metric can be treated as continuous (so that expressions like $m_i - v_i$ are meaningful) and that D_X is approximately symmetric around its median value θ_X. Its null hypothesis is the following [52–54]:

Table 2 Performance of a purely classifier-based pipeline, without the registration of a 3D model

	Mean	SD	Median	Min.	Max.	*p*-value
DSC over scans (%)	95.69	1.87	96.42	89.58	97.49	0.0100
DSC over individual images (%)	95.68	1.84	96.35	89.75	97.42	0.0046
SMAD (mm)	1.12	0.51	0.99	0.68	3.04	0.0103
SMAD, only femoral head (mm)	0.86	0.54	0.67	0.46	3.13	0.0001
Hausdorff (mm)	2.75	1.24	2.44	1.71	7.16	0.0488
Hausdorff, only femoral head (mm)	1.64	0.85	1.43	0.64	4.79	0.0488

A Wilcoxon signed-rank test was performed for each of the six evaluation metrics in order to determine whether the differences with the full pipeline (Table 1) are statistically significant
After the conduction of the six tests, the Holm–Bonferroni correction method was applied and the resulting (corrected) *p*-values are listed in the rightmost column

Table 3 Performance when the geometric feature is not used

	Mean	Std. Dev.	Median	Min.	Max.	*p*-value
DSC over scans (%)	95.80	2.32	96.40	85.04	97.39	0.0007
DSC over individual images (%)	95.78	2.36	96.35	85.29	97.33	0.0008
SMAD (mm)	1.08	0.53	0.98	0.70	3.48	0.0013
SMAD, only femoral head (mm)	0.74	0.33	0.66	0.43	2.07	0.0125
Hausdorff (mm)	2.59	1.42	2.25	1.44	9.08	0.1122
Hausdorff, only femoral head (mm)	1.61	0.71	1.46	1.14	4.87	0.0170

As shown in Table 2, a Wilcoxon signed-rank test was performed for each of the six evaluation metrics to determine whether the differences with the full pipeline (Table 1) are statistically significant
After the conduction of the six tests, the Holm–Bonferroni correction method was applied and the resulting (corrected) *p*-values are listed in the rightmost column

Null hypothesis of WSR : $\theta_X = 0$.

The WSR was chosen over the paired samples *t* test, whose null hypothesis is that the mean value of D_X is zero, because the distributions of the values were found be to non-Gaussian. As a counteraction to the well-known *multiple comparisons problem*, after the conduction of the six tests, the Holm–Bonferroni correction method was applied. The significance level for all the statistical tests was set to 0.05. The resulting (corrected) *p*-values are reported in the rightmost columns in Tables 2 and 3.

Discussion and conclusion

Performance of the proposed pipeline

We performed a validation study on the radial MR scans of 25 patients symptomatic of FAI or AVN. The proposed pipeline yields satisfactory results in terms of both area overlap and border distance to the reference segmentation. As summarized in Table 1, the mean DSC of the segmentation is 96.37%, the mean SMAD is 0.94 mm, and the mean Hausdorff distance is 2.37 mm. There is generally no significant difference between computing the DSC of the whole scan and computing the mean DSC of its images, as the femur occupies more or less the same area in all the images.

In Fig. 6, it can be qualitatively observed that the largest difference with the manual segmentation of a typical case (the one with the median DSC) occurs near the trochanters, whereas there is less difference near the joint space. This agrees with the results in Table 1, as the average SMAD and the average Hausdorff distance are much lower when calculated only on the femoral head (0.64 mm and 1.41 mm, respectively, compared to 0.94 mm and 2.37 mm). We think that this is important for certain applications, since frequently the most critical part of the proximal femur is the one close to the hip joint. It also demonstrates the ability of the proposed method to follow the border of the proximal femur in the regions affected by the crosstalk artifact of the radial scans, which can be challenging during manual segmentation.

Some representative images from six scans are shown in Fig. 7, each exemplifying typical pathological findings and/or challenges for the automated segmentation (three cases where the femoral heads have a highly non-spherical shape and three cases with cam lesions). The border of the automated segmentation is also shown. The pipeline copes with these variations in most cases satisfactory, but mistakes do happen. In Fig. 7, it can be observed that the method did not manage to follow the border of a non-spherical femoral head in a subregion of one case and that a small cam-type lesion was segmented out in another image.

In terms of runtime performance, the proposed pipeline is relatively fast: On a system with a standard Intel i7 CPU dual core at 2.7 GHz, the random forest classification takes about 2.5 s per slice, therefore 35 s for the 14 images of a scan. The subsequent deformable model registration takes approximatively 40 s. The preprocessing stage completes in around 10 s; thus, the pipeline needs around 1.5 min in total to segment the 14 images of a scan and to reconstruct the 3D model of the proximal femur.

Effect of the geometric feature

When only intensity-based features are used, the probability maps frequently have high values for areas which are far away from the femur and the border between background and foreground is not always sharp. These problems are illustrated in Fig. 4b. After the inclusion of the geometric feature, the situation is much improved (Fig. 4c). The effect in performance is quantified in Table 3, which presents the performance of a pipeline that uses only intensity-based features in the classification process: The average, standard deviation, median, minimum and maximum values of all the evaluation metrics get worse, with the average scan DSC decreasing from 96.37 to 95.80%, the average SMAD increasing from 0.94 mm to 1.08 mm and the average Hausdorff distance increasing from 2.37 to 2.59 mm.

This variant was compared with the default pipeline using a WSR test for each evaluation metric, followed by a correction of the p-values with the Holm–Bonferroni method. Within the confidence level of 95%, the difference with the default pipeline with respect to the Hausdorff distance metric was found to not be statistically significant. The difference with respect to each of the remaining five metrics was found to be statistically significant within the confidence level of 95%.

Importance of deformable model registration stage

We also compared the proposed system with a purely classifier-based approach. We found that the latter yields slightly worse mean values of all the six evaluation metrics (Table 2): The average scan DSC decreases from 96.37 to 95.69%; the average SMAD increases from 0.94 to 1.12 mm, and the average Hausdorff distance increases from 2.37 to 2.75 mm. The performance of the classifier-based variant was compared with that of the default pipeline using a WSR test for each evaluation metric, followed by a correction of the p-values with the Holm–Bonferroni method. Within the confidence level of 95%, the difference with the default pipeline with respect to each metric was found to be statistically significant.

In Fig. 8, the difference in performance between the two variants can be assessed visually. This figure presents box-and-whisker plots for the two variants for all the six evaluation metrics. A first observation that we can make from these plots is that there are strong outliers for all the metrics, indicating that their distributions cannot be considered Gaussian. Secondly, the boxes of the plots illustrate the fact that both the median values of the metrics and their distance from the lower quartile (Q_1) to the higher quartile (Q_3), i.e., their interquartile ranges (IQR), are generally better with the deformable model (the only exception is the IQR for the Hausdorff distance, which does not change significantly). Visually, it seems that the difference between the two variants is least significant with respect to the Hausdorff distance in the femur head area and most significant with respect to the SMAD in the femur head area. Both observations agree with the minimum and maximum computed p-values that are reported in Table 2.

We observed that the biggest performance gains with the deformable model registration frequently occur on the scans which are the most challenging to the classifier. For example, the classifier-based method performed worst in the cases with Ids 4, 9 and 12, with average scan DSCs of 90.95%, 89.58% and 93.53% and average SMADs of 2.21 mm, 3.04 mm and 1.60 mm, respectively. The hip joint of case 4 has image findings consistent with Perthes disease, and the cases 9 and 12 correspond to the youngest patients of our dataset (16 years old). Within the variance observed in our dataset, all these three proximal femurs have non-typical appearance. In these cases, the proposed pipeline achieves scan DSCs

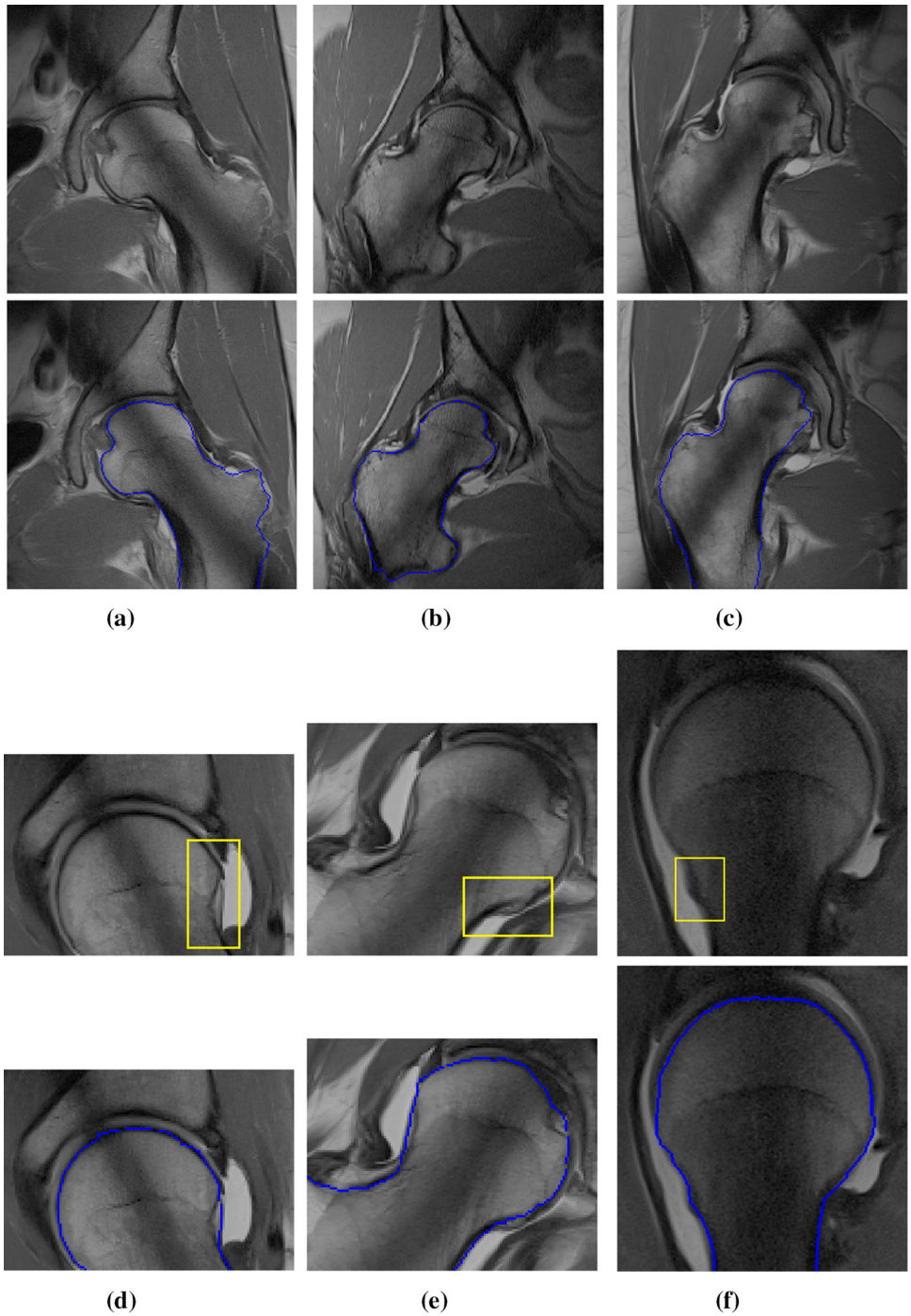

Fig. 7 Visualization of six representative radial images with pathological findings and/or presenting challenges for the segmentation, each corresponding to a different patient. **a**–**c** Non-spherical shape of femur head; **d**–**f** lesions of cam-type FAI (femoral "bumps"). Below every image, the border of the segmentation achieved by the pipeline is also shown. In **c** and **e,** the result of the pipeline has visible mistakes. In **c,** the border of a subregion of the femoral head is not captured correctly and in **e** a lesion is segmented out

90.90%, 95.34% and 95.44% and average SMADs 2.20 mm, 1.39 mm and 1.12 mm for cases 4, 9 and 12, respectively. Our interpretation of these results is that, in some scans, the proximal femur has an appearance that has not been encountered during training, and hence, the classifier has difficulty in capturing its borders accurately. When this happens, the registration of the 3D mesh can help by taking advantage of its explicit modeling of the global shape of the femur.

We can summarize that the advantage of the purely classifier-based approach is that it performs only slightly worse than the complete pipeline while it remains fully automated, as it does not need to be initialized using manually specified landmarks. The full pipeline proceeds with the registration of a deformable model of the proximal femur, which we found to improve the segmentation performance. While it is true that this improvement in performance is small, the resulting registered model of the proximal femur has impor-

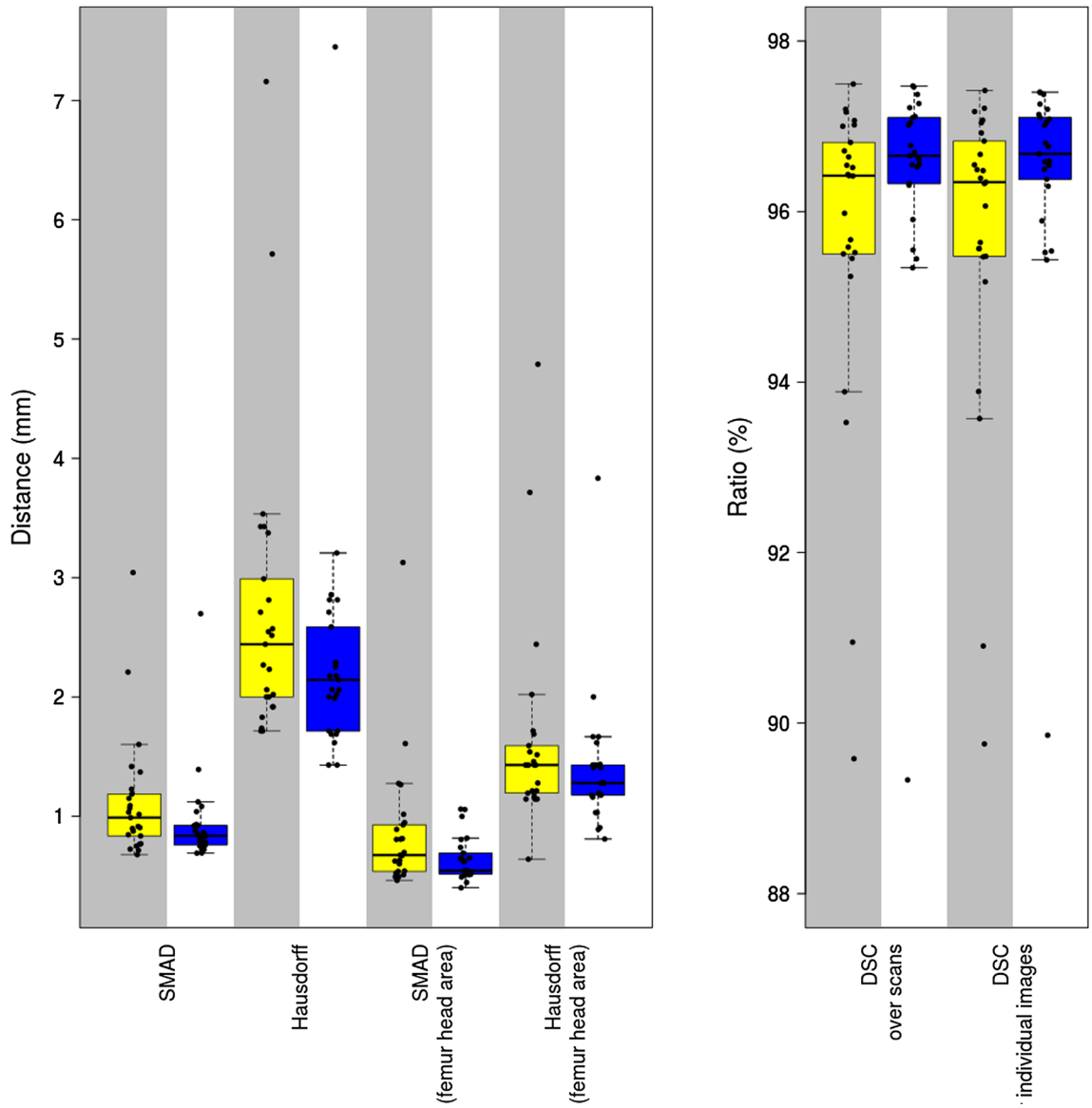

Fig. 8 Box-and-whisker plots of the proposed pipeline and the purely classifier-based variant for all the six evaluation metrics. The proposed pipeline corresponds to the blue boxes (gray background), and the purely classifier-based variant corresponds to the yellow boxes (white background). For clarity of visualization, the data points have been jittered horizontally. The boxes span from the lower quartile (Q_1) to the higher quartile (Q_3), and the band within them corresponds to the median value. Thus, the span of the boxes is equal to the IQR. The two whiskers lay on lowest data point still within 1.5 IQR of the lower quartile and on the highest data point still within 1.5 IQR of the upper quartile

tant implications for clinical applications. Indeed, the output of a classifier is only a binary segmentation mask, whereas the registered model provides rich contextual interpretation of the segmentation result. The latter can be used for the direct localization of anatomical landmarks and for further planning of surgical procedures. For example, it permits the direct computation of clinically relevant morphometric features, such as the head and neck diameters, the sphericity of the head, the length of the proximal femur and the intertrochanteric distance.

Conclusions

We present a pipeline for the segmentation of the proximal femur from radial scans of the hip and the reconstruction of its 3D model. We performed a fivefold cross-validation on a dataset of 25 radial scans of patients symptomatic of FAI or AVN. With respect to a manual, reference segmentation of the proximal femur, the resulting segmentation has an average DSC of 96.37%, an average mean SMAD of 0.94 mm and an average Hausdorff distance of 2.37 mm. In the femoral head

subregion, the average SMAD is 0.64 mm and the average Hausdorff distance is 1.41 mm.

In our view, the main limitation of the presented study is the lack of a comparison with the segmentation achieved from 3D MRI scans. Such an extension of our study is significant, as it will permit the detailed evaluation of the fitted 3D model and the examination of how the segmentation accuracy is affected by the lack of the dense intensity information of the 3D MRI. Unfortunately, we did not have accompanying 3D CT or MRI scans for the cases of our dataset, so such a study was not possible. In the future, we plan to collect additional 3D images, permitting us to further investigate the potential of a method based solely on radial scans to provide a reliable 3D model of the proximal femur.

As part of future work, we also intend to replace the manual picking of the three landmarks for the initialization of the deformable model registration stage with an automated method, thus rendering the whole pipeline fully automated. For this purpose, we could employ a random forest regression localization approach, akin to our previous work [39]. Another possible extension important from a clinical standpoint is the concurrent segmentation of the acetabulum: The critical attribute of FAI is a decreased range of motion; therefore, the modeling of the whole hip joint is a necessary component of a complete system what aims to facilitate the management of the condition. As the acetabulum socket is within the FOV of the radial scans, we expect that the classification stage can be directly extended for this task by simply adding a third class corresponding to the acetabulum. The registration of a 3D model of the acetabulum could be performed with the currently employed method, with the necessary inclusion of a collision detection step that considers both parts of the hip joint.

In conclusion, surgical correction of cam-FAI or AVN is challenging and requires exact preoperative planning. The 3D model of the proximal femur as reconstructed by the proposed method can be utilized for this planning, as it does not require any modification to the MR protocol for these frequent hip diseases. We intend to use this method for diagnosis and planning of surgical correction of cam-FAI or AVN because, unlike 3D MRI, radial MR scans can be included in a routine MRI of the hip.

Funding This study was funded by the Swiss National Science Foundation (Grant number 205321_163224).

Compliance with ethical standards

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individuals included in the study.

References

1. Chughtai M, Piuzzi NS, Khlopas A, Jones LC, Goodman SB, Mont MA (2017) An evidence-based guide to the treatment of osteonecrosis of the femoral head. Bone Jt J 99(10):1267–1279
2. Sullivan JP, Griffith TB, Park CN, Ranawat AS (2017) Advances in 2D and 3D imaging for FAI surgical planning. In: Hip joint restoration. Springer, New York, pp 277–285
3. Leunig M, Beaulé PE, Ganz R (2009) The concept of femoroacetabular impingement: current status and future perspectives. Clin Orthop Relat Res 467(3):616–622
4. Tannast M, Siebenrock KA, Anderson SE (2007) Femoroacetabular impingement: radiographic diagnosis—what the radiologist should know. Am J Roentgenol 188(6):1540–1552
5. Steppacher SD, Huemmer C, Schwab JM, Tannast M, Siebenrock KA (2014) Surgical hip dislocation for treatment of femoroacetabular impingement: factors predicting 5-year survivorship. Clin Orthop Relat Res 472(1):337–348
6. Steppacher SD, Lerch TD, Gharanizadeh K, Liechti EF, Werlen SF, Puls M, Tannast M, Siebenrock KA (2014) Size and shape of the lunate surface in different types of pincer impingement: theoretical implications for surgical therapy. Osteoarthr Cartil 22(7):951–958
7. Lerch TD, Todorski IA, Steppacher SD, Schmaranzer F, Werlen SF, Siebenrock KA, Tannast M (2018) Prevalence of femoral and acetabular version abnormalities in patients with symptomatic hip disease: a controlled study of 538 hips. Am J Sports Med 46(1):122–134. https://doi.org/10.1177/0363546517726983
8. Morita D, Hasegawa Y, Okura T, Osawa Y, Ishiguro N (2017) Long-term outcomes of transtrochanteric rotational osteotomy for non-traumatic osteonecrosis of the femoral head. Bone Jt J 99(2):175–183
9. Petchprapa CN, Dunham KS, Lattanzi R, Recht MP (2013) Demystifying radial imaging of the hip. Radiographics 33(3):E97–E112
10. Chana R, Noorani A, Ashwood N, Chatterji U, Healy J, Baird P (2006) The role of MRI in the diagnosis of proximal femoral fractures in the elderly. Injury 37(2):185–189
11. Cabarrus MC, Ambekar A, Lu Y, Link TM (2008) MRI and CT of insufficiency fractures of the pelvis and the proximal femur. Am J Roentgenol 191(4):995–1001
12. Sutter R, Dietrich TJ, Zingg PO, Pfirrmann CW (2012) How useful is the alpha angle for discriminating between symptomatic patients with cam-type femoroacetabular impingement and asymptomatic volunteers? Radiology 264(2):514–521
13. Klenke FM, Hoffmann DB, Cross BJ, Siebenrock KA (2015) Validation of a standardized mapping system of the hip joint for radial MRA sequencing. Skelet Radiol 44(3):339–343
14. Domayer SE, Mamisch TC, Kress I, Chan J, Kim YJ (2010) Radial dGEMRIC in developmental dysplasia of the hip and in femoroacetabular impingement: preliminary results. Osteoarthr Cartil 18(11):1421–1428
15. Zilkens C, Tiderius CJ, Krauspe R, Bittersohl B (2015) Current knowledge and importance of dGEMRIC techniques in diagnosis of hip joint diseases. Skelet Radiol 44(8):1073–1083

16. Riley GM, McWalter EJ, Stevens KJ, Safran MR, Lattanzi R, Gold GE (2015) MRI of the hip for the evaluation of femoroacetabular impingement; past, present, and future. J Magn Reson Imaging 41(3):558–572
17. Schmaranzer F, Todorski IAS, Lerch TD, Schwab J, Cullmann-Bastian J, Tannast M (2017) Intra-articular lesions: imaging and surgical correlation. In: Seminars in musculoskeletal radiology, vol 21, No. 05. Thieme Medical Publishers, pp 487–506
18. Schmaranzer F, Haefeli PC, Hanke MS, Liechti EF, Werlen SF, Siebenrock KA, Tannast M (2017) How does the dGEMRIC index change after surgical treatment for FAI? A prospective controlled study: preliminary results. Clin Orthop Relat Res 475(4):1080–1099
19. Rathnayaka K, Momot KI, Noser H, Volp A, Schuetz MA, Sahama T, Schmutz B (2012) Quantification of the accuracy of MRI generated 3D models of long bones compared to CT generated 3D models. Med Eng Phys 34(3):357–363
20. Lerch T, Degonda C, Zheng G, Todorski I, Schmaranzer F, Ecker T, Siebenrock K, Tannast M (2017) MR-based 3D PAO planning and simulation of hip impingement is as accurate as CT-based 3D models. German Congress of Orthopedic and Trauma Surgery (DKOU 2017)
21. Xia Y, Fripp J, Chandra SS, Schwarz R, Engstrom C, Crozier S (2013) Automated bone segmentation from large field of view 3D MR images of the hip joint. Phys Med Biol 58(20):7375
22. Schmid J, Kim J, Magnenat-Thalmann N (2011) Robust statistical shape models for MRI bone segmentation in presence of small field of view. Med Image Anal 15(1):155–168
23. Gilles B, Magnenat-Thalmann N (2010) Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. Med Image Anal 14(3):291–302
24. Arezoomand S, Lee WS, Rakhra KS, Beaulé PE (2015) A 3D active model framework for segmentation of proximal femur in MR images. Int J Comput Assist Radiol Surg 10(1):55–66
25. Chandra SS, Xia Y, Engstrom C, Crozier S, Schwarz R, Fripp J (2014) Focused shape models for hip joint segmentation in 3D magnetic resonance images. Med Image Anal 18(3):567–578
26. Zeng G, Yang X, Li J, Yu L, Heng PA, Zheng G (2017) 3D U-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3D MR images. In: International workshop on machine learning in medical imaging. Springer, Cham, pp 274–282
27. Paiement A, Mirmehdi M, Xie X, Hamilton MC (2014) Integrated segmentation and interpolation of sparse data. IEEE Trans Image Process 23(1):110–125
28. Van Assen HC, Danilouchkine MG, Frangi AF, Ordás S, Westenberg JJ, Reiber JH, Lelieveldt BP (2006) SPASM: a 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. Med Image Anal 10(2):286–303
29. Tu Z (2008) Auto-context and its application to high-level vision tasks. In: IEEE conference on computer vision and pattern recognition, CVPR 2008. IEEE, pp 1–8
30. Gao Y, Wang L, Shao Y, Shen D (2014) Learning distance transform for boundary detection and deformable segmentation in ct prostate images. In: International workshop on machine learning in medical imaging. Springer, Cham, pp 93–100
31. Nyúl LG, Udupa JK, Zhang X (2000) New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 19(2):143–150
32. Glocker B, Zikic D, Konukoglu E, Haynor DR, Criminisi A (2013) Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 262–270
33. Criminisi A, Robertson D, Konukoglu E, Shotton J, Pathak S, White S, Siddiqui K (2013) Regression forests for efficient anatomy detection and localization in computed tomography scans. Med Image Anal 17(8):1293–1303
34. Schmid J, Magnenat-Thalmann N (2008) MRI bone segmentation using deformable models and shape priors. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 119–126
35. Volino P, Magnenat-Thalmann N (2000) Implementing fast cloth simulation with collision response. In: Proceedings of the computer graphics international. IEEE, pp 257–266
36. Cootes TF, Hill A, Taylor CJ, Haslam J (1993) The use of active shape models for locating structures in medical images. In: Biennial international conference on information processing in medical imaging. Springer, Berlin, pp 33–47
37. Kraevoy V, Sheffer A (2006) Mean-value geometry encoding. Int J Shape Model 12(01):29–46
38. Kumar S (2003) Discriminative random fields: a discriminative framework for contextual interaction in classification. In: Proceedings of the 9th IEEE international conference on computer vision, 2003. IEEE, pp 1150–1157
39. Chu C, Chen C, Liu L, Zheng G (2015) Facts: fully automatic ct segmentation of a hip joint. Ann Biomed Eng 43(5):1247–1259
40. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128
41. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J (2012) 3D slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 30(9):1323–1341
42. Zikic D, Glocker B, Konukoglu E, Criminisi A, Demiralp C, Shotton J, Thomas OM, Das T, Jena R, Price SJ (2012) Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 369–376
43. Mahapatra D (2014) Analyzing training information from random forests for improved image segmentation. IEEE Trans Image Process 23(4):1504–1512
44. Montillo A, Shotton J, Winn J, Iglesias JE, Metaxas D, Criminisi A (2011) Entangled decision forests and their application for semantic segmentation of CT images. In: Biennial international conference on information processing in medical imaging. Springer, Berlin, pp 184–196
45. Zikic D, Glocker B, Criminisi A (2014) Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. Med Image Anal 18(8):1262–1273
46. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, Ayache N (2011) Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57(2):378–390
47. Louppe G (2014) Understanding random forests: from theory to practice. arXiv Preprint arXiv:1407.7502
48. Criminisi A, Shotton J (eds) (2013) Decision forests for computer vision and medical image analysis. Springer, Berlin
49. Damopoulos D, Glocker B, Zheng G (2017) Automatic localization of the lumbar vertebral landmarks in CT images with context features. In: International workshop and challenge on computational methods and clinical applications in musculoskeletal imaging. Springer, Cham, pp 59–71
50. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
51. Kosov S (2013) Direct graphical models C++ library. http://research.project-10.de/dgm/
52. Woolson RF (2007) Wilcoxon signed-rank test. Wiley, New York, pp 4739–4740. https://doi.org/10.1002/9780471462422.eoct979

53. Li H, Johnson T (2014) Wilcoxon's signed-rank statistic: what null hypothesis and why it matters. Pharmaceutical statistics 13(5):281–285

54. Sheskin DJ (2003) Handbook of parametric and nonparametric statistical procedures. CRC Press, Boca Raton