Medical Image Analysis 52 (2019) 109–118

Contents lists available at ScienceDirect

Medical Image Analysis

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

Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative

Felix Ambellan[a,1,*], Alexander Tack[a,1], Moritz Ehlke[b,a], Stefan Zachow[a,b]

[a] Therapy Planning Group, Zuse Institute Berlin, Berlin, Germany
[b] 1000shapes GmbH, Berlin, Germany

ARTICLE INFO

Article history:
Received 15 August 2018
Revised 6 November 2018
Accepted 16 November 2018
Available online 17 November 2018

Keywords:
Semantic segmentation
Magnetic resonance imaging
Statistical shape models
Deep learning

ABSTRACT

We present a method for the automated segmentation of knee bones and cartilage from magnetic resonance imaging (MRI) that combines *a priori* knowledge of anatomical shape with Convolutional Neural Networks (CNNs). The proposed approach incorporates 3D Statistical Shape Models (SSMs) as well as 2D and 3D CNNs to achieve a robust and accurate segmentation of even highly pathological knee structures. The shape models and neural networks employed are trained using data from the Osteoarthritis Initiative (OAI) and the MICCAI grand challenge "Segmentation of Knee Images 2010" (SKI10), respectively. We evaluate our method on 40 validation and 50 submission datasets from the SKI10 challenge. For the first time, an accuracy equivalent to the inter-observer variability of human readers is achieved in this challenge. Moreover, the quality of the proposed method is thoroughly assessed using various measures for data from the OAI, i.e. 507 manual segmentations of bone and cartilage, and 88 additional manual segmentations of cartilage. Our method yields sub-voxel accuracy for both OAI datasets. We make the 507 manual segmentations as well as our experimental setup publicly available to further aid research in the field of medical image segmentation. In conclusion, combining localized classification via CNNs with statistical anatomical knowledge via SSMs results in a state-of-the-art segmentation method for knee bones and cartilage from MRI data.

1. Introduction

Knee osteoarthritis (OA) is a chronic, degenerative joint disease affecting a significant fraction of the human population (Lawrence et al., 2008). Due to the rising average life expectancy, an increasing obesity, and a prolonged desire for an active lifestyle, research to understand and prevent OA will become even more important. Magnetic Resonance Imaging (MRI) is commonly used to assess knee joint degeneration, especially of the femoral bone (FB), tibial bone (TB), and the respective femoral and tibial cartilage (FC, TC) (Conaghan et al., 2011). Quantitative image-based biomarkers from MRI such as the apparent bone volume divided by total bone tissue volume (Eckstein et al., 2006a) or the volume of the articular cartilage (Eckstein et al., 2006b), show potential for diagnosis of OA, treatment planning, and prognostic purposes. However, clinical studies with a large number of subjects are required to extract quantitative image-based biomarkers indicating pathogenesis of OA and to evaluate the efficacy of therapeutic approaches. Both, the determination of such biomarkers, as well as computer-based surgical planning of interventions affecting the knee require precise segmentations of the respective anatomical structures. Manual segmentation of the knee joint is, however, tedious, subjective, and labor intensive, which renders the analysis of larger cohorts or an individualized therapy planning in clinical routine impractical. Hence, the performance and quality of automated segmentation methods are being constantly improved.

The general development in the field of MRI and Computed tomography (CT) segmentation methods is outlined e.g. in Balafar et al. (2010) and Heimann et al. (2009). In the recent decade especially two effective segmentation approaches have raised attention, Statistical Shape Model (SSM) based segmentation (Heimann and Meinzer, 2009) as well as segmentation methods employing Convolutional Neural Networks (CNNs)

* Corresponding author.
E-mail addresses: ambellan@zib.de (F. Ambellan), tack@zib.de (A. Tack), ehlke@zib.de (M. Ehlke), zachow@zib.de (S. Zachow).
[1] Equal contribution.

https://doi.org/10.1016/j.media.2018.11.009
1361-8415/

(Litjens et al., 2017). The aim of this work is to improve segmentation accuracy of existing approaches even further by combining the strengths of SSM-based and CNN-based methods to render them suitable for individual therapy planning as well as the processing of large cohort data, e.g. the databases of the Osteoarthritis Initiative[2]~(OAI) or the Study of Health in Pomerania[3]~(SHIP). Therefore we propose a method that utilizes 2D CNNs for full resolution images and 3D CNNs for local regions of interest to incorporate information at different levels of detail into the segmentation process, as well as SSMs to support decision making in regions of high uncertainty in the image intensities. The accuracy of our segmentation approach is validated based on a large pool of diverse datasets from the MICCAI challenge "Segmentation of Knee Images 2010" (SKI10)[4] described by Heimann et al. (2010) and from the OAI database. Via utilizing SSMs as anatomical shape prior for regularization and CNNs for learning descriptors of local appearance, our method robustly segments varying MRI sequences, even when the images show subjects with severe OA grades.

We consider the main contributions of this work as follows:

(1) Integration of SSM-based anatomical knowledge in an CNN-based segmentation of FB and TB via a voting scheme.
(2) SSM-based postprocessing for segmentation of knee bones.
(3) 3D CNNs for segmentation of TC and FC.
(4) A thorough assessment of segmentation quality on three different datasets and publication of 507 manual reference segmentations (FB, TB, FC, TC).

2. Related work

Previous methods often employed SSMs (Heimann and Meinzer, 2009) to segment knee bones and cartilage. Vincent et al. (2010) presented a method based on an active appearance model, which was created using a minimum description length approach to optimize correspondences. In the same year, Seim et al. (2010) presented a method that utilizes SSMs for bone segmentation and a multi-object graph optimization for cartilage segmentation. SSM-based methods employ anatomical knowledge via geometric priors, which allows for a robust segmentation even in the presence of artifacts or low image contrast. Such methods usually require heuristically designed models of appearance to adjust the SSM to the image data (Kainmueller, 2014). Often, appearance models are manually taylored to one specific image modality and cannot be easily generalized to cope with differing ones. To alleviate this problem, Cootes et al. (2012) utilized random forest regression voting in 2D which was enhanced to 3D by Norajitra and Maier-Hein (2017). Mukhopadhyay et al. (2016) proposed to derive appearance models directly from the image data via joint dictionary learning.

Meanwhile, CNNs have been employed successfully for medical image segmentation, but only few of these methods address the domain of musculoskeletal research. Prasoon et al. (2013) presented an approach for tibial cartilage segmentation from MRIs using three separate 2D CNNs. Each CNN independently classifies foreground and background pixels from slices in either the axial, coronal, and sagittal image planes of the 3D MRI dataset. In a similar fashion, Liu et al. (2017) applied 2D U-Nets (Ronneberger et al., 2015) as well as the 2D CNN architecture "SegNet" in combination with 3D simplex deformable modeling to obtain 3D segmentations from MRIs. Both methods train 2D convolutional filters from individual slices in the 3D MRI data, since the memory consumption of deep 3D CNNs is often too excessive at the scale of full-resolution 3D medical datasets. Consequently, the image information available to the CNNs is strictly localized and lacks context w.r.t. the surrounding voxel intensities in neighboring slices. This is in contrast to previous SSM-based approaches, where anatomical shape information regularizes the segmentation outcome due to its inherent 3D nature across several neighboring slices or even the whole MRI stack. Ideally, one would train a geometrically constrained CNN, such that an anatomically plausible segmentation result is guaranteed even if the anatomy's boundary cannot be clearly outlined in the image data. However, this is a challenging task and still subject to ongoing research.

A frequently employed strategy (Chen et al., 2014; Christ et al., 2016; Kamnitsas et al., 2017) is to refine the segmentation of a fully convolutional neural network using 3D conditional random fields for postprocessing.

Ravishankar et al. (2017) proposed a method in which the deviation of a U-Net-based binary segmentation to the ground truth is penalized by using a shape regularization network. The encoder-decoder shape regularization network predicts the correct segmentation from a previous (incomplete) one, employing a loss function which considers the resulting mask as well as the encoded representation. A different approach is shape regularization of the CNN-learning-process as proposed by Oktay et al. (2018) through employing an auxiliary network that connects 2D images of objects with 3D binary voxel fields representing the respective objects shape. They introduce shape knowledge during training of CNNs, whereas the method proposed in this work employs shape knowledge explicitly via SSMs during inference.

Tack et al. (2018) showed that 2D CNNs, 3D SSMs, and 3D CNNs can be successfully combined for the segmentation of knee menisci. However, the 2D CNNs and 3D SSMs are merely used as a preprocessing to define a region of interest for the 3D CNNs. SSMs are fitted to the outcome of an initial CNN-based segmentation in order to define regions of interest for a 3D CNN that segments the menisci. The final segmentation outcome via 3D CNNs is not regularized any further.

3. Automated segmentation of bone and cartilage

In this work, we integrate the shape knowledge from SSMs further into the segmentation pipeline. Our aim is to establish an automated method that produces highly accurate segmentations of knee bones and cartilage. The approach should be robust against pathological data, imaging artifacts, as well as the varying image appearance in different MRI sequences. For this purpose, we consecutively apply 2D CNNs, sub-regional 3D CNNs and SSM-based techniques in a segmentation pipeline that operates on knee MRIs as input and generates voxel masks for each segmented structures as output. Application of SSMs is important not only to regularize but also to link both convolutional components in a consistent fashion for the segmentation of bones. Given the margins of the bones, we extract subvolumes along the femoral condyles and tibial plateaus and segment the cartilage in these regions using 3D CNNs.

Fig. 1 depicts the individual stages of the proposed pipeline.

- The first step *CNN-2D* creates initial segmentation masks of FB and TB.
- The second step *SSM adjustment* regularizes the results of step *CNN-2D* by fitting SSMs to these masks.
- The third step *CNN-3D* is a refinement step that employs 3D CNNs to segment small MRI subvolumes at the bone surfaces as given by the preceding *SSM adjustment*.
- The fourth step *SSM postprocessing* uses regions pre-defined on SSMs to enhance the results of *CNN-3D*.

[2] oai.nih.gov.
[3] medizin.uni-greifswald.de/cm/fv/ship.html.
[4] organized by Tobias Heimann and Bram van Ginneken (ski10.org).

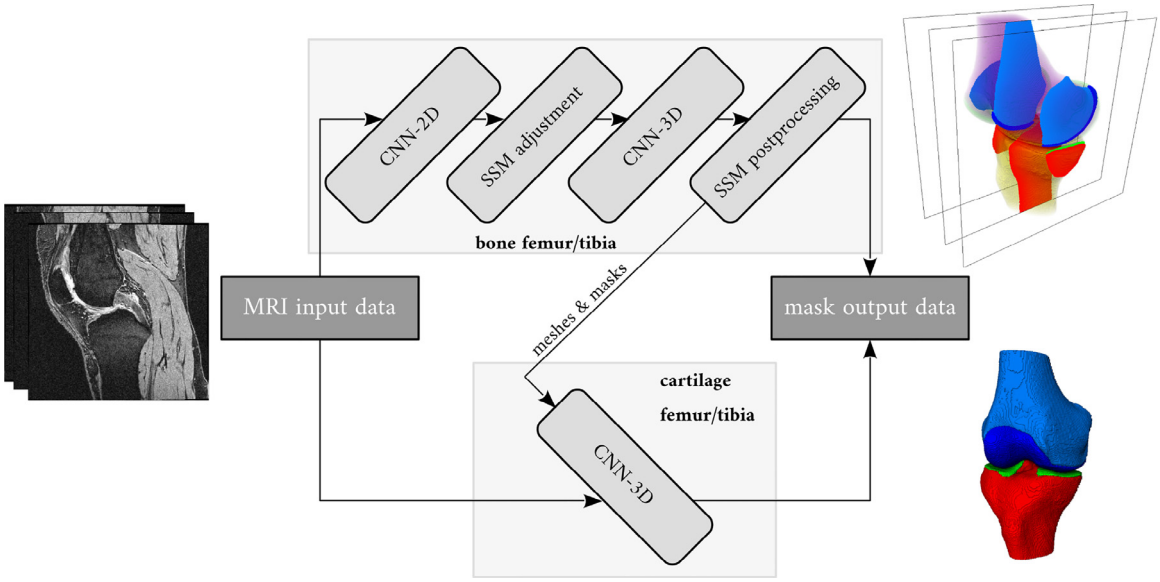

Fig. 1. Proposed pipeline for knee bone and cartilage segmentation. A cascade of CNN and SSM steps yields 3D segmentation masks of the femoral and tibial bone. These masks define the region of interest for a consecutive cartilage segmentation based on 3D CNNs.

- After bone segmentation is finished the FC and TC are segmented using 3D CNNs.

Each step is performed separately for femur and tibia. Thus, CNNs and SSMs are developed independently and individually for both structures.

3.1. CNN-2D

We augment the idea of slice-wise segmentation (Liu et al., 2017) and apply a variant of the 2D U-Net (Fig. 2 left) in order to segment femur and tibia in each slice of the input MRI independently. The number of input channels of the 2D CNN was extended compared to the standard architecture in order to improve spatial consistency of segmentation results between individual slices of the MRI stack. Eight adjacent slices on both sides of the one that is to be segmented are additionally supplied resulting in 17 channels in total. Training is carried out using the (slice-wise) Dice similarity coefficient (Dice, 1945) as a loss function (3). Note that while this method provides additional context information to the CNN, it cannot substitute for true volumetric input as processed by 3D CNNs, since the additional channels are only directly visible to the first convolutional layer. The memory requirements are, however, significantly reduced compared to 3D U-Nets with a similar architecture.

3.2. SSM adjustment

Due to our observation that segmentations of the previous step show inaccuracies in areas of low intensity contrast or imaging artifacts (Fig. 6 left column and Fig. 7 left column) we decided to add an SSM step. The aim of the *SSM adjustment* step is to regularize and to fill holes and notches in the segmentation mask by means of statistical knowledge about the global variation of anatomical shape. For this purpose, an SSM is fitted to the segmentation results from the *CNN-2D* stage. The output is guaranteed to be anatomically plausible (i.e. within the shape span of the SSM) and given as one connected component. Further details on SSMs, their construction and adjustment can be found e.g. in Kainmueller (2014).

The SSM matching procedure is as follows: The side of the knee (left/right) is initially unknown. In order to initialize the SSM considering the side of the knee a template mesh of the condyle region is fitted to the mask, one time affinely and a second time additionally mirrored along the epi-condyle axis. The correct knee side is detected according to the lower distance between template and data of both solutions. Similarity transformation and shape modes of the SSM are adjusted iteratively to fit the vertex positions to the mask:

$$\underset{v(b^{i+1},T^{i+1})}{\arg\min} \left\| \left(v\left(b^i, T^i\right) + \Delta v^i\right) - v\left(b^{i+1}, T^{i+1}\right) \right\|, \quad i \leftarrow i+1, \tag{1}$$

where Δv^i is the displacement along the normals of the vertices resulting from the i-th step, s.t. they are placed as close as possible to the interface between segmentation mask (intensity value = 1) and background (intensity value = 0). The vertex positions, obtained from the SSM w.r.t. shape weights b^i and transformation T^i, are denoted by $v(b^i, T^i)$. The process is iterated until the relative change of the vertex positions is below a certain threshold value or the maximum number of iterations is reached.

3.3. CNN-3D

SSMs, as utilized in the previous stage of the pipeline, cannot express osteophytic details completely (Fig. 6 middle column) since these deformations are highly patient-specific and might not be derived from the training cohort. We approach this issue in the *CNN-3D* step by employing 3D U-Nets (Fig. 2 right) trained on small subvolumes (64x64x16 voxels) localized along the bone contours within the MRI scan. The subvolume's dimensions are chosen by compromising between the input fields size and the memory consumption of the 3D CNNs. This step is carried out utilizing the same 3D architecture, but individual training for every structure to capture anatomical details of bone as well as cartilage tissue. Similar to the 2D-U-Nets, the networks are trained via a loss function based on the Dice similarity coefficient. The loss is, however, defined on volumetric subvolumes in the MRI rather than individual 2D slices, which generally leads to better classification accuracy in local regions of the knee compared to the 2D slicing approach (Fig. 6 left vs. right column).

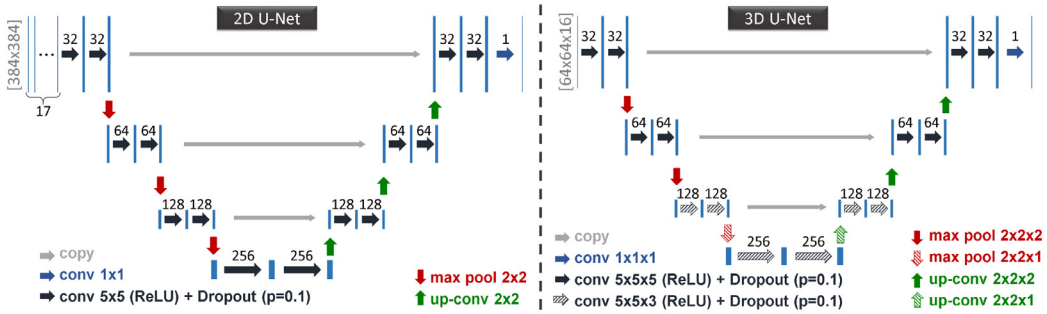

Fig. 2. Architectures of the 2D and 3D U-Nets employed in this work.

To get a feasible number (≈ 1000) of subvolumes they are randomly sampled following a Poisson distribution for the FB and TB segmentation along the outline of the adjusted SSM's mask. The subvolumes for FC segmentation are extracted along the outline of the condyle region in a similar manner. To obtain the subvolumes for TC segmentation, we utilize that the tibial plateaus are almost planar and sample at the superior margin of each sagittal slice in the TB mask, s.t. the in-plane overlap of subvolumes is half its size. A visualization of all sampling procedures is given in Fig. 3.

Since the subvolumens are partially interfering with each other, conflicting labels might be assigned in overlapping regions. A voting scheme solves this ambiguity by summarizing voxel-wise decisions in a voting mask V (2). The outcome of the *SSM adjustment* stage (G) is thereby taken into account, biasing the FB and TB segmentations towards the SSM shape when votes conflict for a voxel. The contribution of the SSM is set to zero for the segmentation of cartilage.

The voting is formularized as follows: Let $\mathbf{x} = (i, j, k)$ be a (global) index triplet of an image voxel. Let further I_s be the mapping from local indices of subvolume B_s to global indices in the image, and let $\mathrm{Im}(I_s)$ be the set of all global indices covered by B_s. The voting mask is computed as:

$$V(\mathbf{x}) = \omega \cdot G(\mathbf{x}) + \sum_{\substack{s \\ \mathbf{x} \in \mathrm{Im}(I_s)}} \left(2 \cdot B_s\left(I_s^{-1}(\mathbf{x})\right) - 1\right), \tag{2}$$

with factor ω set to 25, weighting the contribution of SSM and CNNs roughly equal. The higher ω is chosen, the more emphasis is set on the SSM-based regularization. We have chosen ω considering the density and the expected number of overlaping subvolumes. This choice is based on high trust into the SSM's classification in the sense that almost all subvolumes containing a certain voxel have to agree in order to overrule its decision. This tolerance value is independent of the imaging data and depends on the anatomy of interest only as far as its shape affects the amount of overlapping subvolumes. Finally, a label is assigned to each voxel by means of majority voting, resulting in a segmentation mask U:

$$U(\mathbf{x}) = \begin{cases} 1, & \text{if } V(\mathbf{x}) \geq \tau \\ 0, & \text{if } V(\mathbf{x}) < \tau, \end{cases}$$

where the threshold parameter τ has a fixed value of 1. This way, a simple majority of CNN-classified subvolumes has to up-vote a voxel in order to include it into the final result, if it is not already captured by the SSM. Parameter τ further adjusts the influence of the *CNN-3D* step. The higher its value the more up-votes are needed to classify a voxel as foreground. The choice of τ is indirectly influenced by the anatomy's shape as the number of overlapping subvolumes depends on it.

3.4. SSM postprocessing

The *SSM postprocessing* step finalizes the segmentation of femur and tibia bones. SSM-based postprocessing as utilized in our approach follows the idea to remove false positive voxels from the FB and TB segmentations that are located outside the typical range of osteophytic growth. Voxels are excluded from the segmentation mask depending on their surface distance to the SSM. We found that a conservative regularization after processing the subvolumes with the proposed voting method (*CNN-3D* stage) helps to remove segmentation errors due to the localized nature of the 3D CNN-based classification and low image contrast or noise (Fig. 7, middle and right column).

We identified regions on the SSMs of FB and TB that show higher or lower inter-patient variability in shape (Fig. 4). Regions of higher variability are typically associated with osteophytic growth. To identify these regions in the segmentation, SSMs are matched to the masks obtained from the *CNN-3D* stage and the distance between bone voxel centers and fitted surface is calculated, s.t. every voxel is assigned a distance value either to areas of low or high variability on the surface. This leads to a maximum distance in areas of low variability d_{lv} and to a maximum distance in areas of high variability d_{hv}.

If $d_{hv} \geq d_{lv}$, the postprocessing is terminated since no unexpected deviation is identified. Otherwise, the matched surface is converted to a mask denoted by $\tilde{U}$ and the set difference $D = U \setminus \tilde{U}$ is calculated. D naturally consists of its 3D connected components D_j. Every component is considered as a candidate for removal, if there exists a voxel $x \in D_j$, s.t. its distance to the surface $d_{\mathbf{x}}$ is realized in the area of low variability fulfilling $d_{\mathbf{x}} \geq d_{hv}$ and $d_{\mathbf{x}} \geq tol$. The tolerance was empirically determined as 5.5 mm based on the observation that deviations below this value represent normal morphological variation. Osteophytes and artefacts typically appear with larger distances. Thus this choice is due to anatomical morphology and independent of the imaging modality. For components that fulfill the former rule, every voxel that realizes its distance in the area of low variability is removed from the mask. See Fig. 4 for a schematic depiction.

4. Experiments and results

We evaluate the accuracy of our method on three different publicly available MRI datasets *SKI10, OAI Imorphics* and *OAI ZIB* (Table 1) employing measures that consider the volumetric overlap of segmented structures and the manual segmentations as well as the distances between the respective boundaries.

4.1. MRI datasets

Dataset *SKI10* consists of 60 training, 40 validation, and 50 submission MRIs from the MICCAI SKI10 challenge. All scans were ac-

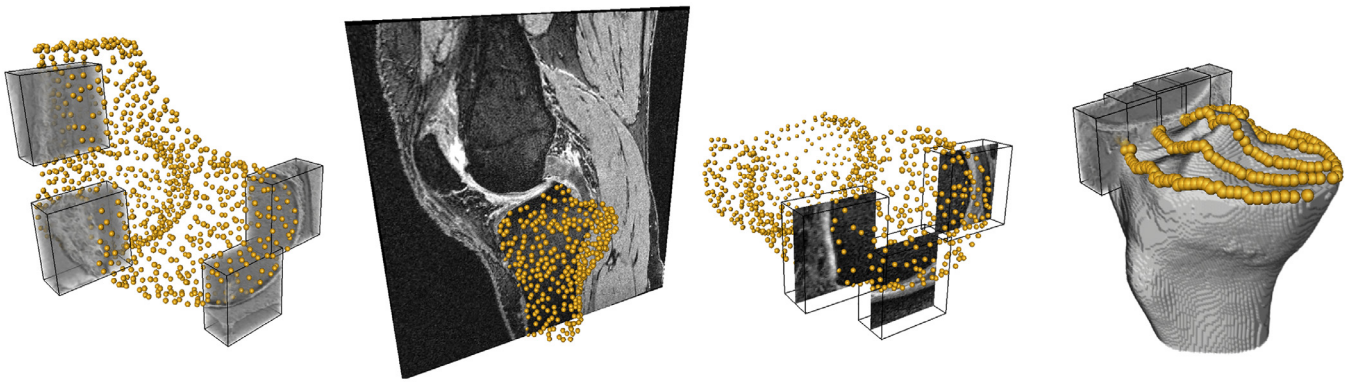

Fig. 3. *CNN-3D* segmentation step: OAI subject ID 9793168. Exemplary sampling of subvolumes. From left to right: FB (1052 points), TB (825 points), FC (845 points) and TC (228 points). Each yellow sphere marks the center of a subvolume.

quired for surgery planning of partial or complete knee replacement, and thus show a high degree of pathological deformations in the knee region.

Dataset *OAI Imorphics* consists of MRI sequences from the OAI database with manual segmentations supplied by Imorphics ($N = 88$). The dataset contains only cases of moderate and severe OA. Dataset *OAI ZIB* consists of additional data from the OAI database for which manual segmentations were carried out thoroughly by experienced users at Zuse Institute Berlin ($N = 507$) starting from automatic segmentations employing (Seim et al., 2010). The data covers the full spectrum of OA grades, with a strong tendency towards severe cases. The *OAI ZIB* segmentations are made publicly available as part of the supplementary material of this manuscript[5]. An overview of the datasets is given in Table 1.

4.2. Experimental setup

The employed SSMs consist of 15,172 vertices and 30,220 faces (FB), and 16,244 vertices and 32,351 faces (TB) independent of the dataset. Construction was done following (Seim et al., 2010). For the *SKI10* dataset, training of CNNs and construction of SSMs is carried out using the 60 training cases. Our method is evaluated for the validation and the submission cases separately.

Two-fold cross-validation studies are performed for datasets *OAI ZIB* and *OAI Imorphics*. For *OAI ZIB*, decomposition is done by random choice (253/254). The resulting groups of 253 and 254 subjects are used for construction of SSMs and for training of CNNs. For *OAI Imorphics* the cohort's subject ids are sorted numerically and split into upper and lower half (44/44). However, since no manual segmentations of bones are available for the *OAI Imorphics* dataset, the SSMs built from the *SKI10* training data are employed. The CNNs are trained using the *OAI Imorphics* baseline data only. Thus, the *OAI Imorphics* 12-month follow-up data (12 m) is exclusively used for evaluation still within the cross-validation setting.

Since the covered variability of SSMs is known to vary with the number of training samples, we perform an experiment to investigate this effect. In this experiment we apply our segmentation method to the *OAI ZIB* dataset using the SSM developed from the *SKI10* training data that consists of 60 shapes instead of 253/254. It is noteworthy that the SSMs employed in this work could be constructed independently from the image data that is to be segmented. The SSMs can be build e.g. from CT scans, since their role is to provide a geometric prior (i.e. a prior on anatomical shape) that is independent of the underlying imaging modality.

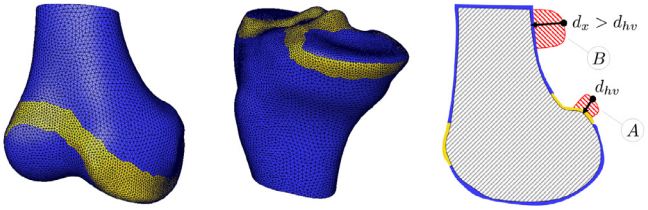

Fig. 4. Predefined regions of high (yellow) and low (blue) variability on SSM surfaces of FB (left) and TB (middle). Right: Schematic visualization of a typical post-processing situation. A large deviation from the SSM (region B) in an area of high confidence in shape (blue) is removed from the segmentation mask.

This means that the SSM related parts of the segmentation pipeline trivially generalize over all imaging modalities.

Implementation of CNNs is done employing Keras (Chollet et al., 2015) using Theano (Al-Rfou et al., 2016) as the backend. The optimization of the CNNs is carried out using the ADAM optimizer (Kingma and Ba, 2014) with a learning rate of 10^{-4} for 2D bone and 3D cartilage as well as 10^{-5} for 3D bone CNNs. All calculations regarding SSMs are carried out using *Amira ZIB Edition 42.2017*[6] (Stalling et al., 2005).

4.3. Measures of segmentation accuracy

The accuracy of our method is evaluated using the Dice Similarity Coefficient (DSC), average surface distance (ASD), root mean square surface distance (RSD), maximum surface distance (MSD), volume difference (VD), and volume overlap error (VOE). All these measures are symmetric apart from VD that is considered relative to the manual segmentation. Volumetric measures (3), (4), (5) are suitable for assessing the segmentation results globally. However, volume-based measures provide limited sensitivity to errors on the boundaries of the segmentation if the segmented volume is relatively large. Thus, small features, such as osteophytes that are important e.g. for diagnostic purposes are not adequately represented by these measures alone. We therefore also include surface distance measures (6), (7), (8) in the evaluation, which are sensitive to segmentation errors on the anatomical boundary. Within the following formulas, A denotes the set of manually segmented (ground-truth) voxels and B denotes the segmentation result from the automated method; ∂A and ∂B represent the boundary of A and B. The boundary contains every voxel with at least one neigh-

[5] doi.org/10.12752/4.ATEZ.1.0.

[6] amira.zib.de/download.html.

Table 1
Summary of the datasets used for training and validation. Images were acquired either once per patient (baseline) or twice with an additional 12-month follow-up (12 m).

	SKI10	OAI Imorphics	OAI ZIB
MRI scanner	GE, Siemens, Philips, Toshiba, Hitachi. Mostly 1.5T, some 3T, a few 1T	Siemens 3T Trio	Siemens 3T Trio
MRI sequence	Many (T1, T2, GRE, Spoiled-GRE) partly with fat suppression	DESS	DESS
Acquisition plane	sagittal	sagittal	sagittal
Image resolution [mm]	0.39 × 0.39 × 1.0	0.36 × 0.36 × 0.7	0.36 × 0.36 × 0.7
Manual segmentations	bones and cartilage	cartilage	bones and cartilage
Number of subjects	60 training 40 validation 50 submission	88	507
Sex (male,female)	n.a.	(45,43)	(262,245)
Age [years]	n.a.	61.24 ± 9.98	61.87 ± 9.33
BMI [kg/m^2]	n.a.	31.06 ± 4.61	29.27 ± 4.52
rOA grade (0,1,2,3,4)	n.a.	(0,0,15,56,17)	(60,77,61,151,158)
timepoints	baseline	baseline, 12 m	baseline

Table 2
Segmentation accuracy for the *SKI10* validation dataset. Results for FB, TB are reported for different steps of our segmentation pipeline: Step one (*CNN-2D*), step three (*CNN-3D*), and for the complete pipeline including the postprocessing (Full).

		ASD (mm)	RSD (mm)	VD (%)	VOE (%)
CNN-2D	**FB**	0.47 ± 0.23	0.90 ± 0.48	–	–
	TB	0.68 ± 1.58	1.36 ± 3.28	–	–
CNN-3D	**FB**	0.43 ± 0.13	0.75 ± 0.28	–	–
	TB	0.37 ± 0.11	0.63 ± 0.26	–	–
Full	**FB**	0.43 ± 0.13	0.74 ± 0.27	–	–
	TB	0.35 ± 0.07	0.59 ± 0.19	–	–
	FC	–	–	7.18 ± 10.51	20.99 ± 5.08
	TC	–	–	4.29 ± 12.34	19.06 ± 5.18

Total score as computed employing the SKI10 metrics:
71.2 ± 11.2 (CNN-2D), 73.6 ± 7.6 (CNN-3D), 74.0 ± 7.7 (Full).

bor that is not part of the respective segmentation mask. The number of voxels on the boundary ∂A, ∂B is written as $n_{\partial A}$, $n_{\partial B}$. Lastly $|\cdot|$ denotes a volume and $\|\cdot\|_2$ the usual Euclidean norm.

$$\mathrm{DSC} = 100 \cdot \frac{2\,|\,B \cap A\,|}{|\,B\,| + |\,A\,|} \tag{3}$$

$$\mathrm{VOE} = 100 \cdot \left(1 - \frac{\mathrm{DSC}}{200 - \mathrm{DSC}}\right) \tag{4}$$

$$\mathrm{VD} = 100 \cdot \frac{|\,B\,| - |\,A\,|}{|\,A\,|} \tag{5}$$

$$\mathrm{ASD} = \frac{1}{n_{\partial A} + n_{\partial B}} \left(\sum_{i=1}^{n_{\partial A}} \min_{b \in \partial B} \|a_i - b\|_2 + \sum_{j=1}^{n_{\partial B}} \min_{a \in \partial A} \|b_j - a\|_2 \right) \tag{6}$$

$$\mathrm{RSD} = \sqrt{\frac{1}{n_{\partial A} + n_{\partial B}} \left(\sum_{i=1}^{n_{\partial A}} \min_{b \in \partial B} \|a_i - b\|_2^2 + \sum_{j=1}^{n_{\partial B}} \min_{a \in \partial A} \|b_j - a\|_2^2 \right)} \tag{7}$$

$$\mathrm{MSD} = \max \left(\max_{a \in \partial A} \min_{b \in \partial B} \|a - b\|_2,\ \max_{b \in \partial B} \min_{a \in \partial A} \|b - a\|_2 \right) \tag{8}$$

4.4. Results

Table 2 summarizes our results for the *SKI10* validation dataset. Results are reported for the bones after step one (CNN-2D), step three (CNN-3D), as well as for the complete segmentation pipeline (Full). Our proposed method reaches a total score of 74.0 ± 7.7 in terms of the SKI10 metrics (Heimann et al., 2010). This is a notable improvement w.r.t. scores reported in previous works (Table 3). Moreover, our method achieves a total score of 75.73 on the *SKI10* submit data and is currently ranked first as of November 2018[7].

Table 3
SKI10 validation data: Our approach (bold) yields the best of all published results as of November 2018.

Imorphics (Vincent et al., 2010)	ZIB (2010) (Seim et al., 2010)	BioMedIA (Wang et al., 2013)
52.3 ± 8.6	54.4 ± 8.8	56.5 ± 9.2
Liu et al. (Liu et al., 2017)	**Biomediq (Dam et al., 2015)**	**ZIB (2018)**
64.1 ± 9.5	67.1 ± 8.0	**74.0 ± 7.7**

The results for the *OAI Imorphics* dataset are shown in Table 4. For FC the DSC is 89.4% for baseline and 89.1% for 12 m. For medial tibial cartilage (MTC) and lateral tibial cartilage (LTC) the DSC is 86.1% resp. 90.4% for baseline, and 85.8% resp. 90.0% for 12 m. The ASD is smaller than the image resolution (< 0.36 mm) for both, FC and TC. Fig. 5 compares the results obtained from the method proposed in this work to the results reported by Dam et al. (2015). For the *OAI Imorphics* dataset, our method shows an improvement in segmentation accuracy by approx. 5 percentage points for FC, approx. 5 percentage points for MTC, and approx. 4 percentage points for LTC w.r.t. the DSC.

The accuracy evaluation for the *OAI ZIB* dataset is summarized in Table 5. The DSC is 98.6% for FB, 98.5% for TB, 89.9% for FC, and 85.6% for TC. Again, the ASD is smaller than the image resolution for bone as well as for cartilage (< 0.36 mm). Additionally, the results of an ablation study are shown as well in Table 5 including the accuracy employing the *SKI10* SSM, containing less training shapes than the *OAI ZIB* ones. The usage of a smaller SSM is not changing the magnitude of the results for any distance measure. However, performing t-tests between the results of the *OAI ZIB* and the *SKI10* SSM, the SSM containing more training data yields significantly better ($P < 0.0001$) results for all measures but the MSD (tibial bone, $P = 0.08$).

[7] ski10.org/results.php.

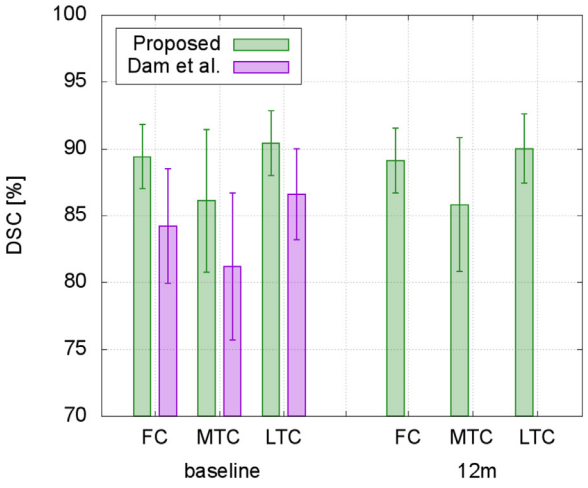

Fig. 5. *OAI Imorphics* dataset: Our results compared to those reported in Dam et al. (2015) w.r.t. Dice similarity coefficient (DSC). The evaluation of their method was done for baseline timepoint only.

Computation for the whole segmentation pipeline is measured as 9 m 22 s (exemplarily) on a consumer-grade workstation (CPU: Intel Xeon E5-2650 v3, 2.30GHz; GPU: GeForce GTX 980 Ti).

5. Discussion

The presented method was evaluated thoroughly using datasets from the SKI10 challenge as well as from the OAI database. Accuracy was evaluated using volume-based and distance-based measures to provide a sound analysis w.r.t. global and local level of detail. The proposed method consistently achieved a high segmentation accuracy, despite severely arthritic knees and various different MRI sequences. For the first time, a total score greater than 75 was reached on the *SKI10* submission data, which is comparable to the inter-observer variability of two expert readers (Heimann et al., 2010). Moreover, better results were achieved for the *SKI10* validation data compared to the total scores reported by previous methods. Although our score was below 75 for this dataset, a relative improvement of approx. 10% was achieved compared to the previously best result reported by Dam et al. (2015) (74.0 vs. 67.1). They also gave results for the segmentation of FC as well as TC of the *OAI Imorphics* dataset (Fig. 5). For the baseline timepoint, the 3D CNN-based method proposed in this work achieves a higher segmentation accuracy. We believe this is possibly due to the application of deep learning methods instead of the feature detectors used previously.

In two ablation studies we have shown the influence of each proposed step of our segmentation pipeline (see Table 2 and Table 5). Especially for the *SKI10* validation data the increase of the total score with each step is noticeable (71.2 vs. 73.6 vs. 74.0). For the *OAI ZIB* dataset the clearest trend within the accuracy measures can be observed for the maximum error. Each step of our pipeline is decreasing the mean error as well as the standard deviation. We evaluated the influence of the SSM's training data size on our segmentation results and found that for SSMs of different size a similar magnitude of quality is reached. However, the results differed significantly in the statistical sense between the smaller *SKI10* and the *OAI ZIB* SSMs containing more training data. It was thus shown that more variability covered in the employed SSMs leads to better segmentation accuracy. This is a typical behavior working with SSMs in segmentation (Lamecker et al., 2004) and as such an expected phenomenon. Moreover, the *SKI10* data was acquired using different MRI scanners and different MRI sequences compared to the *OAI ZIB* data. We note that to employ an SSM for the task of knee bone segmentation within the presented method, it is not necessary to build it from geometries reconstructed of the target imaging modality, since SSMs rely on ge-

Table 4
Cartilage segmentation accuracy for the *OAI Imorphics* dataset at baseline (00 m) and 12-month follow-up (12 m).

		DSC (%)	ASD (mm)	RSD (mm)	MSD (mm)	VD (%)	VOE (%)
00 m	**FC**	89.4 ± 2.41	0.19 ± 0.08	0.50 ± 0.30	6.65 ± 2.99	7.0 ± 6.04	19.1 ± 3.88
	MTC	86.1 ± 5.33	0.26 ± 0.23	0.63 ± 0.55	5.16 ± 2.93	8.0 ± 17.15	24.1 ± 7.74
	LTC	90.4 ± 2.42	0.17 ± 0.06	0.41 ± 0.16	3.93 ± 2.07	6.9 ± 7.14	17.5 ± 3.96
12 m	**FC**	89.1 ± 2.41	0.20 ± 0.09	0.53 ± 0.33	6.86 ± 3.16	7.6 ± 6.78	19.6 ± 3.86
	MTC	85.8 ± 5.00	0.28 ± 0.22	0.67 ± 0.55	5.25 ± 3.06	6.6 ± 16.38	24.5 ± 7.37
	LTC	90.0 ± 2.57	0.18 ± 0.06	0.44 ± 0.19	4.08 ± 2.11	7.2 ± 7.74	18.1 ± 4.16

Table 5
Bone and cartilage segmentation accuracy for the *OAI ZIB* dataset. Results are reported for different steps of our segmentation pipeline: Step one (CNN-2D), step three (CNN-3D), and for the complete pipeline including the postprocessing (Full). Additionally, results are shown employing the *SKI10* SSM containing 60 training shapes instead of the *OAI ZIB* SSM, which was constructed using 253/254 training shapes, depending on the cross-validation group.

		DSC (%)	ASD (mm)	RSD (mm)	MSD (mm)	VD (%)	VOE (%)
Full	**FB**	98.6 ± 0.30	0.17 ± 0.05	0.35 ± 0.09	2.93 ± 1.24	-0.09 ± 0.87	2.8 ± 0.58
	TB	98.5 ± 0.33	0.18 ± 0.06	0.37 ± 0.18	3.16 ± 2.03	-0.03 ± 0.82	2.9 ± 0.63
	FC	89.9 ± 3.60	0.16 ± 0.07	0.38 ± 0.17	5.35 ± 2.50	1.5 ± 5.87	18.1 ± 5.90
	TC	85.6 ± 4.54	0.23 ± 0.12	0.60 ± 0.38	6.35 ± 4.36	-1.0 ± 11.92	24.9 ± 6.79
CNN-3D	**FB**	98.4 ± 0.44	0.18 ± 0.05	0.35 ± 0.11	3.02 ± 1.57	0.01 ± 0.89	2.8 ± 0.61
	TB	98.3 ± 0.44	0.19 ± 0.09	0.41 ± 0.33	3.45 ± 2.80	0.01 ± 0.93	3.0 ± 0.75
CNN-2D	**FB**	98.6 ± 0.36	0.18 ± 0.07	0.40 ± 0.22	4.17 ± 2.93	0.07 ± 0.87	2.7 ± 0.69
	TB	98.6 ± 0.40	0.17 ± 0.08	0.38 ± 0.22	3.58 ± 2.44	-0.07 ± 1.00	2.7 ± 0.76
SKI10 SSM	**FB**	98.4 ± 0.31	0.19 ± 0.05	0.38 ± 0.09	3.10 ± 1.28	-0.04 ± 0.89	3.1 ± 0.60
	TB	98.4 ± 0.35	0.20 ± 0.07	0.39 ± 0.19	3.27 ± 2.16	-0.03 ± 0.85	3.1 ± 0.66

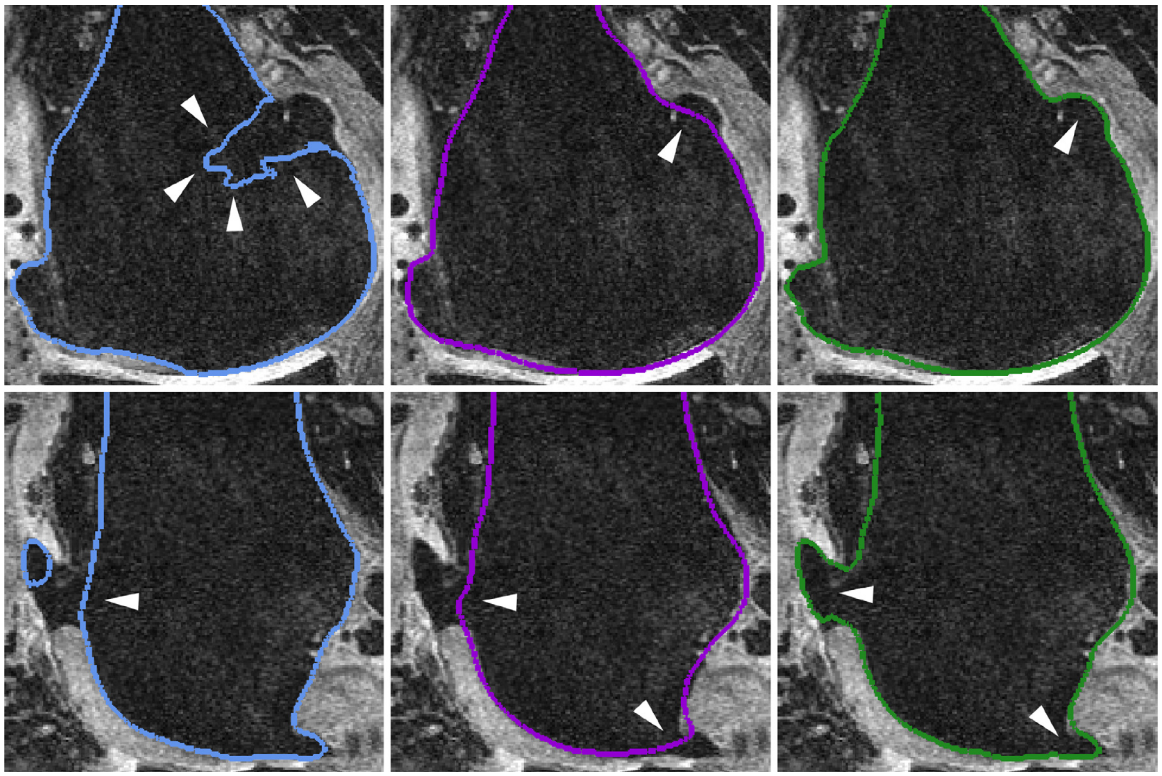

Fig. 6. Segmentation of osteophytic regions in different stages (colored contours). *CNN-2D* stage (left) is error prone, *SSM adjustment* (middle) smoothly regularizes and *CNN-3D* (right) segments osteophytes precisely.

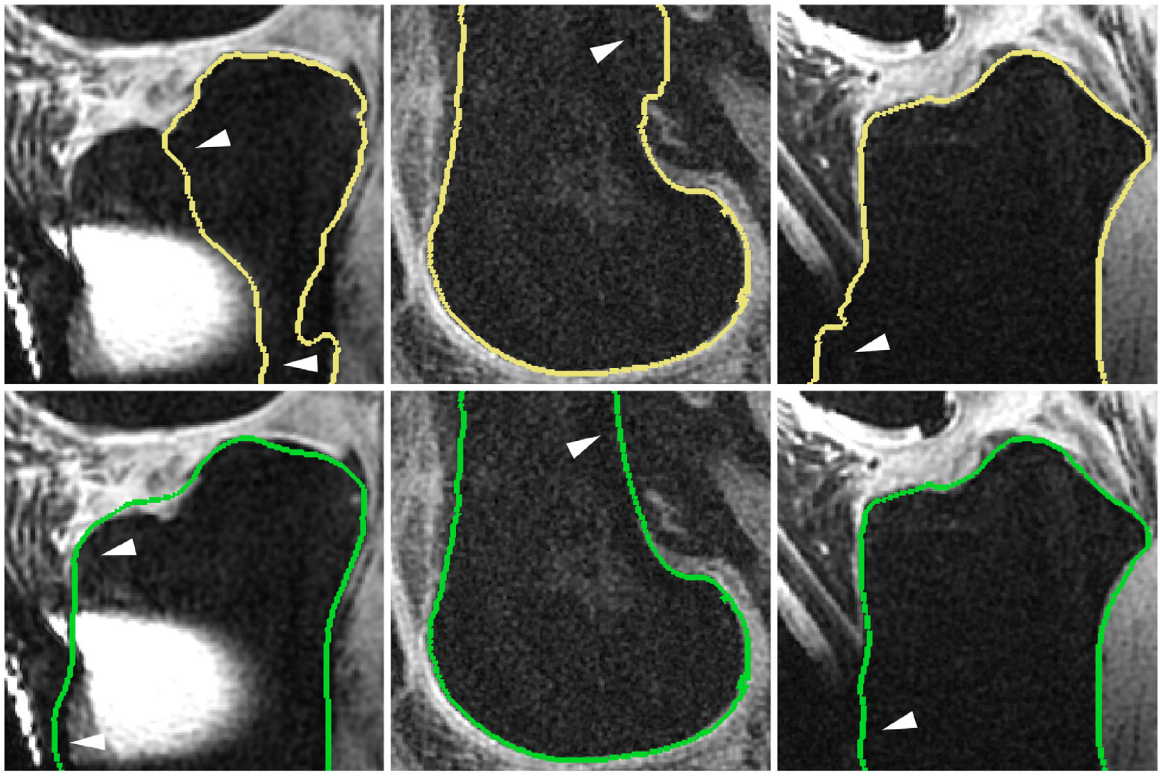

Fig. 7. Left column: An image artifact results in an erroneous *CNN-2D* classification (top). An anatomically plausible segmentation (bottom) is achieved employing *SSM adjustment* regularization. Middle and right column: *CNN-3D* segmentation introduces errors in the shaft region due to insufficient image contrast (top). *SSM postprocessing* corrects in an anatomically plausible manner (bottom).

ometrical information only and are not connected to MRI images in any way.

An interesting option could be to change the type of SSM used within the second step of the pipeline, since other kinds of SSMs will give different modes of variation for the same input data. Recently, von Tycowicz et al. (2018) presented a Riemannian shape model based on differential coordinates that has proven to be more sensitive to OA related pathological changes in bony tissue. Such non-linear model could improve the SSM fit in step two and therefore lead to an increased overall segmentation accuracy especially for osteoarthritic knees. The adjustment algorithm itself of the SSM fit could potentially be improved by consideration of anisotropic covariances, as proposed by Bernard et al. (2017).

Within the *SSM postprocessing* step of our method, specific regions are defined manually. In order to make the proposed segmentation approach applicable for other anatomies as well applying a learning technique to characterize these trust regions for the SSM is desirable. Moreover, an evaluation of 3D conditional random fields can be performed in this context (Kamnitsas et al., 2017).

In our experience, the automated method reduces the time effort for an accurate segmentation of knee bones and cartilage at least by a factor of six compared to manual segmentations by an experienced reader (>1h). However, large scale databases for studying the OA disease, such as provided by the OAI, can contain 50.000 or more MRIs. Using our implementation, it would take 43 weeks to segment the full OAI database on a singe computational node. We therefore aim at reducing the computational time of the algorithm further as well as distributing the workload over several nodes. Our goal is to segment the full OAI database and to make the results publicly available in the near future.

6. Conclusion and future work

We presented a new method which achieved for the first time a segmentation accuracy as good as human experts on the SKI10 dataset by combining CNNs with statistical shape knowledge. The results prove that explicit anatomical information notably augments the segmentation process and that a close incorporation between CNNs and shape knowledge is desirable.

Thus, a promising line of future work is to investigate approaches that couple CNNs and SSMs more directly, e.g. by introducing learned appearance from CNNs to an SSM segmentation framework. One way of combining CNNs and SSMs could be to solve a well known problem: The manual design of appearance models as described e.g. in Kainmüller et al. (2007), which are driving the SSM to a given MRI dataset. These appearance models are usually evaluated for each vertex of an SSM and can be learnt for a given population as shown by Norajitra and Maier-Hein (2017). Instead of random regression forests a CNN could be used, which then decides for each vertex, to which target position it should be adjusted in order to minimize the distance between the SSM and the ground truth object boundary.

Interesting as well is an investigation whether a connection can be established between the method of Oktay et al. (2018) and ours. The coefficients of the SSM-shape-representation could probably be connected to a voxel representation of that shape in order to augment it or maybe even to replace it within their setting. This could lead to an approach that includes shape knowledge during training *and* inference. A rather different focus is set in Maron et al. (2017), who developed a technically sound method employing a convolution operator directly on surfaces and thus allowing for learning geometrical features of certain shapes with CNNs. Their approach is interesting as shape based regularization although some research has to be done in order to incorporate it into a CNN segmentation framework.

We chose the dimensions of the subvolumes used as input for the 3D CNNs in a compromise between size and memory consumption. Recently, Heinrich et al. (2018) proposed a very memory efficient solution for 3D image segmentation. This approach could be employed to utilize larger subvolumes or even the full image for segmentation of knee bones and cartilage, which could consecutively be refined by shape knowledge.

Apart from that, shape information might be added to the CNN-training-procedure by integrating the surface distance to the ground-truth into the loss function.

In the future, a more direct (end-to-end) combination of SSMs and CNNs might preserve the benefits of shape knowledge and offers potential of CNN training for diverse and low-scale datasets. To aid this development the manual segmentations created by experienced users at the Zuse Institute Berlin are made publicly available as part of this publication. These segmentation masks (as well as our supplied cross-validation setting) can be utilized for training of new methods and for subsequent benchmarking.

Acknowledgements

We would like to thank Heiko Ramm (né Seim, 1000shapes GmbH) for valuable insights into SSM-based segmentation methods. We also thank Irene Ziska, Agnieszka Putyra, and Robert Joachimsky for revising automated presegmentations of the *OAI ZIB* dataset. The authors gratefully acknowledge the financial support by the German federal ministry of education and research (BMBF) research network on musculoskeletal diseases, grant no. 01EC1408B (Overload/PrevOP) and grant no. 01EC1406E (TOKMIS). The Osteoarthritis Initiative is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

References

Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Bleecher Snyder, J., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P.-L., Cho, K., Chorowski, J., Christiano, P., Cooijmans, T., Côté, M.-A., Côté, M., Courville, A., N. Dauphin, Y., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M, Dumoulin, V., Ebrahimi Kahou, S., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I., Graham, M., Gulcehre, C., Hamel, P., Harlouchet, I., Heng, J.-P., Hidasi, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, Ć, Lee, S., Lefrancois, S., Lemieux, S., Léonard, N., Lin, Z., Livezey, J.A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P.-A., Mastropietro, O., McGibbon, R.T., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J., Schulman, J., Schwartz, G., Vlad Serban, I., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Ramana Subramanyam, S., Sygnowski, J., Tanguay, J., van Tulder, G., Turian, J., Urban, S., Vincent, P., Visin, F., de Vries, H., Warde-Farley, D., Webb, D.J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S., Zhang, Y., 2016. Theano: A python framework for fast computation of mathematical expressions. arXiv http://arxiv.org/abs/1605.02688

Balafar, M.A., Ramli, A.R., Saripan, M.I., Mashohor, S., 2010. Review of brain mri image segmentation methods. Artif. Intell. Rev. 33 (3), 261–274.

Bernard, F., Salamanca, L., Thunberg, J., Tack, A., Jentsch, D., Lamecker, H., Zachow, S., Hertel, F., Goncalves, J., Gemmar, P., 2017. Shape-aware surface reconstruction from sparse 3d point-clouds. Med. Image Anal. 38, 77–89.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv:1412.7062.

Chollet, F., et al., 2015. Keras. https://keras.io.

Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M., et al., 2016. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 415–423.

Conaghan, P., Hunter, D., Maillefert, J., Reichmann, W., Losina, E., 2011. Summary and recommendations of the oarsi fda osteoarthritis assessment of structural change working group. Osteoarthritis Cartilage 19 (5), 606–610.

Cootes, T.F., Ionita, M.C., Lindner, C., Sauer, P., 2012. Robust and accurate shape model fitting using random forest regression voting. In: European Conference on Computer Vision, pp. 278–291.

Dam, E.B., Lillholm, M., Marques, J., Nielsen, M., 2015. Automatic segmentation of high-and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. J. Med. Imaging 2 (2), 024001.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.

Eckstein, F., Burstein, D., Link, T.M., 2006. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed. 19 (7), 822–854.

Eckstein, F., Cicuttini, F., Raynauld, J.P., Waterton, J.C., Peterfy, C., 2006. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. Osteoarthritis Cartilage 14, 46–75.

Heimann, T., Meinzer, H.-P., 2009. Statistical shape models for 3d medical image segmentation: a review. Med. Image Anal. 13 (4), 543–563.

Heimann, T., Morrison, B.J., Styner, M.A., Niethammer, M., Warfield, S.K., 2010. Segmentation of knee images: a grand challenge. In: MICCAI Workshop on Medical Image Analysis for the Clinic, pp. 207–214.

Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al., 2009. Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE Trans. Med. Imaging 28 (8), 1251–1265.

Heinrich, M.P., Oktay, O., Bouteldja, N., 2018. Obelisk-one kernel to solve nearly everything: Unified 3d binary convolutions for image analysis. International conference on Medical Imaging with Deep Learning.

Kainmueller, D., 2014. Deformable meshes for medical image segmentation: accurate automatic segmentation of anatomical structures. Springer.

Kainmüller, D., Lange, T., Lamecker, H., 2007. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In: Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, pp. 109–116.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv:1412.6980.

Lamecker, H., Lange, T., Seebass, M., 2004. Segmentation of the Liver using a 3D Statistical Shape Model. Technical Report. ZIB, Takustr. 7, 14195 Berlin.

Lawrence, R.C., Felson, D.T., Helmick, C.G., Arnold, L.M., Choi, H., Deyo, R.A., Gabriel, S., Hirsch, R., Hochberg, M.C., Hunder, G.G., et al., 2008. Estimates of the prevalence of arthritis and other rheumatic conditions in the united states: part ii. Arthritis Rheumatol. 58 (1), 26–35.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, F., Zhou, Z., Jang, H., Samsonov, A., Zhao, G., Kijowski, R., 2017. Deep convolutional neural network and 3d deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn. Reson. Med. 79 (4), 2379–2391.

Maron, H., Galun, M., Aigerman, N., Trope, M., Dym, N., Yumer, E., Kim, V.G., Lipman, Y., 2017. Convolutional neural networks on surfaces via seamless toric covers. ACM Trans. Graph. 36 (4), 71.

Mukhopadhyay, A., Victoria, O.S.M., Zachow, S., Lamecker, H., 2016. Robust and accurate appearance models based on joint dictionary learning data from the osteoarthritis initiative. In: International Workshop on Patch-based Techniques in Medical Imaging, pp. 25–33.

Norajitra, T., Maier-Hein, K.H., 2017. 3D statistical shape models incorporating landmark-wise random regression forests for omni-directional landmark detection. IEEE Trans. Med. Imaging 36 (1), 155–168.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., de Marvao, A., Dawes, T., O'Regan, D.P., et al., 2018. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging 37 (2), 384–395.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E.B., Nielsen, M., 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 246–253.

Ravishankar, H., Venkataramani, R., Thiruvenkadam, S., Sudhakar, P., Vaidya, V., 2017. Learning and incorporating shape models for semantic segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 203–211.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241.

Seim, H., Kainmueller, D., Lamecker, H., Bindernagel, M., Malinowski, J., Zachow, S., 2010. Model-based auto-segmentation of knee bones and cartilage in MRI data. In: MICCAI Workshop Medical Image Analysis for the Clinic, pp. 215–223.

Stalling, D., Westerhoff, M., Hege, H.-C., 2005. Amira: a highly interactive system for visual data analysis. In: The Visualization Handbook, pp. 749–767.

Tack, A., Mukhopadhyay, A., Zachow, S., 2018. Knee menisci segmentation using convolutional neural networks: data from the osteoarthritis initiative. Osteoarthritis Cartilage 26 (5), 680–688.

von Tycowicz, C., Ambellan, F., Mukhopadhyay, A., Zachow, S., 2018. An efficient riemannian statistical shape model using differential coordinates: with application to the classification of data from the osteoarthritis initiative. Med. Image Anal. 43, 1–9.

Vincent, G., Wolstenholme, C., Scott, I., Bowes, M., 2010. Fully automatic segmentation of the knee joint using active appearance models. In: MICCAI Workshop Medical Image Analysis for the Clinic, pp. 224–230.

Wang, Z., Donoghue, C., Rueckert, D., 2013. Patch-based segmentation without registration: application to knee MRI. In: MICCAI Workshop Machine Learning in Medical Imaging, pp. 98–105.