Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging

Fang Liu[1,*], **Zhaoye Zhou**[2], **Hyungseok Jang**[1], **Alexey Samsonov**[1], **Gengyan Zhao**[1], and **Richard Kijowski**[1]

[1]Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA.

[2]Department of Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota, USA.

Abstract

Purpose: To describe and evaluate a new fully automated musculoskeletal tissue segmentation method using deep convolutional neural network (CNN) and three-dimensional (3D) simplex deformable modeling to improve the accuracy and efficiency of cartilage and bone segmentation within the knee joint.

Methods: A fully automated segmentation pipeline was built by combining a semantic segmentation CNN and 3D simplex deformable modeling. A CNN technique called SegNet was applied as the core of the segmentation method to perform high resolution pixel-wise multi-class tissue classification. The 3D simplex deformable modeling refined the output from SegNet to preserve the overall shape and maintain a desirable smooth surface for musculoskeletal structure. The fully automated segmentation method was tested using a publicly available knee image data set to compare with currently used state-of-the-art segmentation methods. The fully automated method was also evaluated on two different data sets, which include morphological and quantitative MR images with different tissue contrasts.

Results: The proposed fully automated segmentation method provided good segmentation performance with segmentation accuracy superior to most of state-of-the-art methods in the publicly available knee image data set. The method also demonstrated versatile segmentation performance on both morphological and quantitative musculoskeletal MR images with different tissue contrasts and spatial resolutions.

Conclusion: The study demonstrates that the combined CNN and 3D deformable modeling approach is useful for performing rapid and accurate cartilage and bone segmentation within the knee joint. The CNN has promising potential applications in musculoskeletal imaging.

---

*Correspondence to: Fang Liu, PhD, Department of Radiology, Wisconsin Institutes for Medical Research, 1111 Highland Avenue, Madison, WI 53705-2275, USA. Tel: 612-222-9728; Fax: 608-263-5112; fliu37@wisc.edu.

INTRODUCTION

Osteoarthritis is one of the most prevalent chronic diseases worldwide with the knee being one of the joints most commonly affected by the disease (1–3). Developing sensitive imaging methods has been considered crucial to understanding the pathogenesis of osteoarthritis and measuring the efficacy of disease modifying therapies. MR imaging is one of the most commonly used methods for longitudinal assessment of joint degeneration in osteoarthritis research studies. The technique is sensitive for detecting early cartilage degeneration (4) and can be used to evaluate other joint structures that can be sources of pain in patients with osteoarthritis (5). MR images can also be used to create three-dimensional (3D) geometric maps of the knee joint showing measurements of cartilage thickness at each point on the articular surface that can provide quantitative measures to assess disease-related and treatment-related changes in cartilage over time (6,7).

To obtain quantitative measures of cartilage thickness of the knee joint from MR images, the bone-cartilage interface and cartilage surface boundary needs to be segmented over the entire articulating surface. Traditionally, this is done by having users scroll through stacks of two-dimensional (2D) images and manually segmenting cartilage from adjacent joint structures. However, manual segmentation is extremely time-consuming, generally requiring several hours for complete cartilage segmentation (8), and its efficiency and repeatability is influenced by the level of human expertise (9). Therefore, there has been recent interest in developing fully automated techniques for segmenting the knee joint structures (10).

The majority of recently proposed methods for fully automated cartilage segmentation can be categorized into model-based and atlas-based approaches. Model-based approaches apply a prior statistical shape model of image object and try to best match the model to the target image. Fripp et al. (11) developed a fully automated boundary-based cartilage segmentation method using the active shape model built from a training set of triangulated surface representations of each of the bone regions in the knee. Seim et al. (12) proposed an approach combining 3D statistical shape modeling and generalized Hough transform to segment bone and cartilage. The bone is first segmented using statistical shape modeling and cartilage is subsequently reconstructed with a multi-object technique using prior knowledge of cartilage thickness. Vincent et al. (13) developed a method to use 3D active appearance model for cartilage segmentation. This method reconstructs 3D surface using a minimum description length approach to group-wise image registration. On the other hand, in atlas-based approaches, one or multiple references are generated by aligning and merging manually segmented images into specific atlas coordinate spaces. Tamez-Pena et al. (14) presented a feasibility study to segment knee cartilage using an atlas database built from MR images of many healthy subjects. Dam et al. (15) evaluated a method combining multi-atlas registration with voxel classification in a multi-structure setting using K nearest neighbor algorithm. Despite promising results by model-based and atlas-based cartilage segmentation methods, both approaches perform poorly when there is high subject variability and significant differences of local features. In addition, both approaches rely on a priori knowledge of knee shapes and require high computation cost and relatively long segmentation times. There are also various semi-automated techniques that have been used for musculoskeletal image segmentation using "region growing" (16), "live wire" (17),

"edge detection" (18), and "active contour" (19) methods. Although these semi-automated methods can achieve good results, they generally require a great deal of user interaction and therefore can be time-consuming and effort-demanding when compared to fully automated methods. The difficulty of segmenting musculoskeletal structures is primarily attributed to factors including varying image contrast, partial volume effects, inherent heterogeneity of image intensity, image artifacts as a result of motion and field inhomogeneity, and variability of joint structures. Therefore, developing an accurate and robust musculoskeletal segmentation method is challenging and still remains an unmet need (7,20).

In recent years, tremendous progress has been made for natural image processing and analysis as a result of the revival of deep convolutional neural networks (CNN). Adaptation of CNN deep learning methods to improve disease detection and image analysis has been shown to have significant impact on medical imaging (21). More specifically, there have been several recent successful applications of CNN deep learning methods for medical image segmentation. Pereira et al. (22) developed a CNN with small convolutional kernels and relatively deep structure for segmenting brain tumors. Brosch et al. (23) proposed a deep 3D CNN method combining a convolutional encoder and a deconvolutional network with shortcut connection for segmenting brain lesions of subjects suffering with multiple sclerosis. Moeskops et al. (24) designed an algorithm to segment brain tissues with patch-wise training and multiple pathways estimation. A recent study by Avendi et al. (25) described a combined CNN network and 2D deformable model for segmenting left ventricle on cardiac MR images. A study by Kamnitsas et al. (26) applied a CNN for segmenting brain lesions in MR images where 3D convolutional layers and 3D fully connected conditional random field (CRF) were used for improving performance. Despite rapid advance of deep learning techniques, few previous studies have described CNN applications in musculoskeletal imaging. One study using CNN to quantify osteoarthritis severity on knee radiographs demonstrated superior results over currently used methods (27). Another recent pilot study published by Prasoon et al. (28) used multi-plane 2D CNN networks to classify tibia cartilage voxels that outperformed a 3D method. CNN techniques have clearly shown promising potentials in musculoskeletal imaging.

Segmenting multiple structures in musculoskeletal imaging is essentially a multi-class labeling problem related to semantic segmentation that is extensively studied using CNN techniques in recent years. A widely used fully convolutional network semantic segmentation method was first proposed by Long et al. (29) to directly apply convolutional CNN to the whole image for spatially dense prediction. This method provides high spatial accuracy using skip connections to extract features from multiple network layers at computing final segmentation result. Among all the methods, the convolutional encoder-decoder (CED) network that consists of a paired encoder and decoder networks is favorable in multiple studies (30,31). Ronneberger et al. (30) developed U-Net that is closely related to fully convolutional network and has a CED structure with skip connections between encoder and decoder. The U-Net decoder transfers entire feature maps from encoder to decoder and concentrate them to obtain up-sampled feature maps via deconvolution. This network improved the state-of-the-art methods in segmenting neuronal structures in electron microscopy. Badrinarayanan et al. (31) proposed a CED network called SegNet built on the deep structure of VGG16 network (32). In contrast to other CED methods, a unique up-

sampling approach in decoder was introduced to use pre-stored max-pooling indices. This network offers a simple yet efficient alternative to deconvolution for recovering high resolution image features and achieved top performance in multiple semantic segmentation contests.

In this study, we propose to improve musculoskeletal cartilage and bone segmentation by using a fully automated method combining a semantic segmentation CED network and 3D simplex deformable modeling (33). The SegNet has been selected as the core of this fully automated segmentation method based on three advantages. First, SegNet is designed specifically to carry out pixel-wise semantic labelling for high resolution images. High resolution segmentation is critical to preserve fine detail in musculoskeletal imaging, for instance, to accurately delineate thin cartilage from adjacent bone in subjects with osteoarthritis. Second, SegNet uses an efficient up-sampling scheme in decoder that possesses high memory and computation efficiency during inference. This advantage could potentially provide rapid segmentation solution for various applications in musculoskeletal imaging. Finally, SegNet is easy to implement and has support from fast parallel computing method by Caffe framework (34) that improves segmentation efficiency and allows use of the method for multiple musculoskeletal applications.

The purpose of this study is to describe and evaluate the pilot application of CNN deep learning method for segmentation of musculoskeletal tissues from MR images and to build a framework to pipeline this fully automated process. The new segmentation method is compared with other state-of-the-art techniques using a publicly available knee MR image data set from the Segmentation of Knee Images 2010 (SKI10) challenge (20). The study also demonstrates the ability of the new technique to segment musculoskeletal tissues on morphologic and quantitative MR images with different tissue contrast using smaller image training data sets. We show that the new CNN method improves the accuracy and efficiency of musculoskeletal tissue segmentation when compared to currently used techniques.

THEORY

Convolutional Encoder-Decoder Architecture

SegNet was selected as a core CED segmentation engine for constructing the deep learning network. The SegNet architecture was first introduced for rapid and efficient road scene understanding applications and benchmarked on outdoor and indoor scene data sets for multiple class segmentation tasks (31). An illustration of the SegNet network is shown in Figure 1. This network is composed of an encoder network and a decoder network. The encoder network is essentially the same as the 13 convolutional layers of the popular VGG16 network (32) designed for object recognition. Each encoder layer group in the encoder networks consists of 2D convolution with a set of filters, batch normalization (BN) (35), and rectified-linear unit (ReLU) activation (36) followed by a max-pooling process. The encoder layer group is repeated five times to achieve efficient data compression while probing robust and spatial invariant image features.

To recover high resolution feature maps using up-sampling, the SegNet network removes the fully connected layers that are typically used in other CNN segmentation networks after the

encoder network. Instead, a decoder network with the same number of five decoder layer groups is applied following the encoder network output. The decoder network is a reverse process of the encoder and consists of an up-sampling layer followed by a trainable convolutional layer, BN, and ReLU activation. Specifically, the up-sampling layer performs non-linear up-sampling by initializing an up-sampled feature map with zeroes and only filling feature values at the pre-stored pooling indices that are computed from the corresponding max-pooling layer of the encoder (31). Because it only requires storing the pooling indices instead of whole feature maps of the encoder, as required by architectures that use skip connections for up-sampling (29,30), this up-sampling process provides efficient preservation for high resolution image boundaries and features at low cost of memory consumption in the testing phase. The final layer of decoder network is a multi-class soft-max classifier (34) that produces class probabilities for each individual pixel at the same image resolution as the input image.

Three-Dimensional Deformable Model

Based on the class probability output from the final soft-max classifier from SegNet, the label for each voxel is defined as the class index with the highest class probability. The class labels of whole 3D image volume is used as the initial tissue classification, and the tissue boundary is extracted by using marching cube algorithm (37) for each individual segmentation objects including the femur and tibia and overlying femoral and tibial cartilage. The segmentation is then refined through 3D simplex deformable modeling to preserve the smooth boundary and maintain overall object shape. The simplex model uses simplex mesh that is robust and efficient in permitting smooth deformation. In general, the simplex deformation can be considered as a problem to solve vertices (V) motion of all simplex meshes at a Newtonian law of motion given as (33)

$$m\frac{d^2V}{dt^2} = -\lambda\frac{dV}{dt} + F_{in} + F_{ex}, \quad [1]$$

where m is the vertex mass, λ is the damping factor ranging from 0 to 1 and is selected to tradeoff between deformation efficiency and stability, F_{in} is the internal force from the simplex mesh network to ensure continuity, and F_{ex} is the external force from constraints to regularize distance between mesh and image boundaries. To obtain the image boundaries for a given image volume, the input image is first denoised by using Perona-Malik anisotropic diffusion filter (38) and then normalized by using voxel-wise sigmoid transformation. The image gradient is calculated by applying a recursive Gaussian filter (39) on the normalized image and then set as image boundary for estimating external force. Equation [1] can be solved to achieve a numerical solution using the central finite differences (40) with an explicit scheme at a discrete time point (t) as (33)

$$V^{t+1} = V^t + (1-\lambda)\left(V^t - V^{t-1}\right) + \alpha F_{in} + \beta F_{ex}, \quad [2]$$

where α and β are tunable internal and external force factors. This numerical solution is conditionally stable only if α and β are within a certain interval to generate a convergence (i.e., a successful deformable result). The α is found experimentally to ensure stable iterative deformation with a value below 0.5. The β is typically selected to be a value smaller than 1.0 (33). Detailed explanation of 3D simplex deformable modeling (33) is beyond the scope of this paper. Deformable refinement is carried out to delineate each individual joint structure including the femoral, tibial, and patellar bone and cartilage, and then the deformed results are combined for final tissue segmentation in a 3D class label volume.

Full Segmentation Algorithm

The full segmentation algorithm is summarized in Figure 2. The 3D image volume is dissembled as a stack of 2D images. In the training phase, the SegNet uses a batch of 2D image slices as input and compares them to the corresponding pixel-wise class labels in the training data set. An objective function using cross-entropy loss (29) is estimated for the training loss and the SegNet network parameters are updated until the optimization reaches a predefined maximum iteration step. The network weights to use at testing time are selected based on the iteration wherein the training loss is the lowest among all the iterations. In the testing phase, the well-trained SegNet is used as a front-end segmentation classifier to segment test 2D images. The 3D pixel-wise tissue labels from the SegNet output are passed to iterative processing filter to fill holes and remove small isolated objects using a connected-component filter. The processed labels are then sent to a process to generate 3D simplex mesh by using marching cube algorithm. The simplex mesh is then sent to the 3D simplex deformable process with each individual segmentation objects being separately refined based on the source images. The final 3D segmentation is generated by combining all the deformed objects. Voxels falling into multiple tissues at the final stage of combination after deformable process are simply assigned to the tissue with the highest class probability from the SegNet output.

The segmentation algorithm is implemented in Python language. We use an implementation of SegNet based on Caffe that supports graphics processing unit computation (34). The image processing filters and 3D simplex deformable subroutine is implemented using C + + language and libraries from the Insight Segmentation and Registration Toolkit (ITK) and the Visualization Toolkit (VTK) from KitWare Inc.

METHODS

Image Data Sets

The cartilage and bone segmentation method was evaluated on both a publicly available SKI10 knee image data set and clinical knee image data sets acquired from patients at our institute. The former application was carried out to provide benchmark data to compare with other currently used state-of-the-art segmentation methods. The later application was carried out to demonstrate the applicability and usefulness of the new segmentation method using images with different tissue contrast and spatial resolution and smaller training data sets.

SKI10 Challenge Knee Image Data Set

The Segmentation of Knee Images 2010 (SKI10, www.ski10.org) competition is one of the featured challenges in the Grand Challenge series of workshops hosted by the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in 2010 (20). The idea of this competition is to provide a set of open knee image data sets and standardized evaluation measures to allow comparison of segmentation results from different research groups. The image data sets in SKI10 consists of 100 sagittal T_1-weighted spoiled gradient-echo (SPGR) knee images (60 for training and 40 for testing) acquired at 1T, 1.5T, and 3.0T using scanners from all major MR vendors. The images were acquired with an in-plane resolution of 0.4 mm $\times$ 0.4 mm, approximately 300×360 matrix (actual matrix dimension varies with the size of the knee joint), a slice thickness of 1 mm, and approximately 100 image slices. A multi-class mask was manually segmented by experienced experts for each image data set with the following values: 0 = background, 1 = femur, 2 = femoral cartilage, 3 = tibia, and 4 = tibial cartilage.

Clinical Knee Image Data Sets

The study was carried out in compliance with Health Insurance Portability and Accountability Act of 1996 regulations, with approval from our Institutional Review Board, and with all subjects signing informed consent. Two image data sets were acquired on patients undergoing a clinical MR examination of the knee at our institution using a 3T scanner (Discovery MR750, GE Healthcare, Waukesha, Wisconsin, USA) and 8-channel phased-array extremity coil.

The first image data set consisted of a morphologic sagittal frequency selective fat-suppressed 3D fast spin-echo (3D-FSE) sequence acquired on 60 patients with TR/TE = 2216/23.6 ms, 16 cm field of view, 384×384 matrix, 2-mm slice thickness, and 45 image slices. The second image data set consisted of quantitative multi-echo spin-echo T_2 mapping sequence acquired on 100 patients with a TR = 1500 ms, TE = 7, 16, 25, 34, 43, and 52 ms, 90° flip angle, 16 cm field of view, 256×192 matrix (reconstructed to 256×256 matrix), 3-mm slice thickness, 1-mm slice gap, and 18 image slices. T_2 maps were created using a previously described mono-exponential fitting method (41). Manual cartilage segmentation on the T_2 maps using a previously described direct T_2 segmentation technique (DST) method (42) and 3D-FSE images was carried out by a research assistant with 8 years of segmentation experience under the supervision of a fellowship-trained musculoskeletal radiologist. A multi-class mask was created for each image data set with the following values: 0 = background, 1 = femur, 2 = tibia, and 3 = patella.

Network Training and Deformable Processing

All the input 2D images were normalized using local contrast normalization (43) and resampled to 320×320 using bilinear interpolation before sending to SegNet. When training SegNet, the encoder and decoder network weights were initialized using the initialization scheme of He et al. (44) and updated using stochastic gradient descent (45) with a fixed learning rate of 0.01 and momentum of 0.9. An increased number of feature maps from 64 to 512 were generated by convolutional layers from the first layer group to the last in encoder network with the same number of feature maps applied to paired convolutional layers in

decoder network. The max-pooling filter used a 2×2 window and stride 2 that leads to sub-sampled feature maps with a size reduction factor of 2. The network was trained using the cross-entropy loss (29) as an objective function where the loss is a summation of all pixel loss in a mini-batch of four slices (limited by graphics processing unit global memory size). To achieve training loss convergence, network training was carried out with 60,000, 40,000, and 20,000 iteration steps that corresponds to 22, 88, and 74 epochs for the SKI10 images, 3D-FSE images, and T_2 maps, respectively. The training data was shuffled before each epoch and each mini-batch of four slices is chosen in order, which ensure all images were processed once during every single epoch. In addition, a three-fold cross-validation was carried out on the 3D-FSE images and T_2 maps, respectively, for training and testing the network.

The 3D simplex deformable modeling was carried out for each individual object of interest following the initial segmentation result from SegNet. The parameters for the deformable process were empirically selected and included an internal force factor α of 0.3, external force factor β of 0.01, and damping factor λ of 0.65 (Eqs. [1] and [2]). A total 50 iteration steps were used to ensure surface smoothness and to maintain overall object shape.

To analyze SegNet and compare the performance with other networks, the U-Net network was implemented in the segmentation pipeline by simply replacing the SegNet with the U-Net at default network settings proposed in the original U-Net paper (30) and without data augmentation.

All training and testing were carried out on a desktop computer hosting an Intel Xeon W3520 quad-core CPU with 12 GB DDR3 RAM and a Nvidia Quadro K4200 graphic card (1344 CUDA cores, 4GB GDDR5 RAM) and running a 64-bit Linux operation system.

Evaluation of Segmentation Accuracy

Several quantitative metrics originally proposed in the SKI10 challenges (20) were used to evaluate the accuracy of the new segmentation method on the different image data sets. No manual correction of the segmented cartilage and bone contours was carried out. To evaluate the accuracy of bone segmentation, the average symmetric surface distance and root mean square symmetric surface distance were used. The average symmetric surface distance was defined as the average surface distance between each boundary voxels in the segmentation and the closest boundary voxels in the manual reference in millimeter with the smaller number indicating better overlap between segmentation and reference. The root-mean-square symmetric surface distance takes the square of each distance value and then calculates the root of the average value. To evaluate the accuracy of cartilage segmentation, the volumetric overlap error (VOE) and volu-metric difference (VD) were used. The VOE was defined as

$$\mathrm{VOE} = 100 \left(1 - \frac{S \cap R}{S \cup R} \right), \quad [3]$$

where S and R are the set of voxels in segmentation and manual reference respectively. The smaller VOE value indicated more accurate segmentation with l00% value indicating that there was no overlap at all between the segmentation and manual reference. The VD was defined as

$$\text{VD} = 100\frac{S - R}{R}. \quad [4]$$

Because the segmentation and manual reference does not necessarily need to be overlapped to generate smaller VD values, this metric was mainly used to indicate the size difference of the segmented cartilage. The VOE and VD values were calculated within a region-of-interest (ROI) that comprised the cartilage areas of greatest interest in most osteoarthritis research studies including the central tibial plateau, central femoral condyle, and trochlea. The ROI data for the SKI10 challenge was provided along with the image data set from the official website (www.ski10.org). To allow comparison of segmentation accuracy between the different methods used in the SKI10 competition, a combined single score (ranging from 0 to 100 points) based on the abovementioned quantitative metrics was estimated by using a scoring system provided by the SKI10 organization committee (20). A higher score indicated more accurate segmentation. To account for inter-observer variation, an independent second rater carried out manual segmentation and the outcome for each metric corresponded to 75 points. The score for a testing algorithm linearly decreased as the error increased comparing with the second rater's error. The overall score was taken by averaging scores for all metrics of all images in the data set. For the clinical 3D-FSE and T_2 map image data sets, a large ROI region was drawn in each of the central three consecutive slices to delineate the medial and lateral tibial plateau, medial and lateral femoral condyles, and patella. For comparison of the segmented T_2 maps, a paired Student's t-test was also used to evaluate difference between the average T_2 relaxation time of the femoral, tibial, and patellar cartilage measured using contours created with the manual and fully automated segmentation methods.

RESULTS

Table 1 shows the average numbers of the quantitative accuracy metrics for bone and cartilage segmentation for the 40 test patients in the SKI10 image data set. A score of 64.1 was calculated for our segmentation method using the SKI10 standard scoring system and was compared with the top six segmentation methods that had submitted estimated segmentations for the 40 test patients (www.ski10.org). Our segmentation method using SegNet ranked the second top overall performance among all state-of-the-art cartilage and bone segmentation methods. Our method using SegNet achieved the top performance for femur segmentation and the second top performance for tibia, femoral cartilage, and tibial cartilage segmentation. The U-Net with deformable modeling underperformed the SegNet but also achieved comparable results in comparison to other state-of-the-art methods. Apparently, the 3D deformable modeling improved the final segmentation performance from the output of both SegNet and U-Net.

An example of the output class probability map is demonstrated in Figure 3 and was derived from the final multi-class soft-max layer of the decoder network of Seg-Net. Figure 3a shows a sagittal fat-suppressed T_1-weighted SPGR source image from a patient in the SKI10 image data set. Figures 3b–f are pixel-wise probability maps for background, femur, femoral cartilage, tibia, and tibial cartilage, respectively, which is indexed from 0 to4. Figure 3g is the pixel-wise labeling result obtained from SegNet by combining Figures 3b–f where each individual pixel is assigned with a class index from which the class probability is the highest.

Figure 4 shows an example of a patient in the SKI10 image data set with severe cartilage loss. Figures 4a and 4b show a sagittal fat-suppressed T_1-weighted SPGR source image through the medial knee joint along with the manually delineated tissue labels. The cartilage ROIs for evaluating segmentation accuracy are enclosed by green and red lines for the femoral and tibial cartilage respectively in Figure 4a. The segmentation output of this single slice from SegNet is shown in Figure 4c. The segmentation result from SegNet demonstrates good agreement with the overall contours of the ground truth. However, the segmentation does not create a smooth tissue boundary for cartilage and bone that is indicated by arrows in Figure 4c. The 3D knee joint model directly from SegNet in Figure 4d also demonstrates irregular cartilage surface and stair-like boundary on the bone surface that is primarily caused by low resolution in the slice direction. Figures 4e and 4f show the denoised input image after Perona-Malik anisotropic diffusion filter and the tissue boundary gradient map after recursive Gaussian filter, respectively, which provide metrics for calculating the external force in 3D simplex deformable modeling. Figure 4g is the final result of our segmentation method along with the whole knee joint 3D rendering in Figure 4h. The final segmentation provides excellent agreement with the overall shape of the ground truth as well as maintaining desirable smooth cartilage and bone surfaces indicated by arrows in Figure 4g. For this case, there was a segmentation accuracy of average symmetric surface distance 0.47 mm and 0.46 mm and root-mean-square symmetric surface distance 0.97 mm and 1.11 mm for the femur and tibia, respectively, indicating small deviation from the manually segmented ground truth. There was a segmentation error of VOE20.7% and 19.9% and VD 6.5% and −1.6% for femoral and tibial cartilage, respectively, indicating accurate cartilage segmentation.

Another example of a subject with severe cartilage loss and a bone edema lesion in the SKI10 image data set is demonstrated in Figure 5. Figures 5a and 5b show a sagittal fat-suppressed T_1-weighted SPGR source image through the medial knee joint along with the manually delineated tissue labels. As shown in Figure 5c, the segmentation result from SegNet demonstrates good agreement with the overall contours of the ground truth, but there are mislabeled bone regions (dashed arrow) as a result of increased bone signal intensity caused by the bone marrow edema lesion. The final segmentation result corrects mislabeled bone regions and also demonstrates a smooth cartilage and bone surface as indicated by arrows in Figure 5d. In addition, the 3D rendering of the final segmentation in Figure 5f is superior to the direct result from SegNet in Figure 5e, and both segmentations demonstrate good estimation of cartilage and bone structure. For this case, there was a segmentation accuracy of average symmetric surface distance 0.62 mm and 0.53 mm and root-mean-square symmetric surface distance1.12 mm and 1.11 mm for the femur and tibia,

respectively, and a segmentation error of VOE 28.3% and 28.8% and VD 18.8% and −9.1% for femoral and tibial cartilage, respectively.

The averaged segmentation error of VOE and VD is shown in Table 2 for SegNet and U-Net with and without 3D deformable modeling. The SegNet outperformed U-Net in both the 3D-FSE and T_2 mapping image data sets. The segmentation with 3D deformable modeling outperformed the ones without deformable modeling in both the 3D-FSE and T_2 mapping data sets.

Figure 6 shows an example of cartilage segmentation carried out using the combined SegNet and 3D simplex deformable segmentation method in a clinical patient imaged with a fat-suppressed 3D-FSE sequence that has different tissue contrast than the fat-suppressed T_1-weighted SPGR sequence used in the SKI10 challenge. Figure 6c shows the fully automated segmented femoral, tibial, and patellar cartilage on one sagittal image slice that demonstrates good agreement with the manually segmented cartilage shown in Figure 6b. The 3D cartilage rendering created using the fully automated segmentation method is shown in Figure 6d.

Figure 7 shows an example of cartilage segmentation carried out using the combined SegNet and 3D simplex deformable segmentation method in a clinical patient imaged with a T_2 mapping sequence. The top row in Figure 7 shows T_2 maps from several selected sagittal slices from the medial, central, and lateral knee joint. There are high T_2 values for bone marrow and subcutaneous fat, intermediate values for cartilage and muscle, and low values for meniscus, tendon and ligament. The middle and bottom rows demonstrate the segmented femoral, tibial, and patellar cartilage from the T_2 maps using the manual segmentation and the automated segmentation methods, respectively. The average T_2 relaxation time of femoral, tibial, and patellar cartilage was 39.8 ms, 36.4 ms, and 32.0 ms for the manual segmentation and 38.4 ms, 37.2 ms, and 33.1 ms for the fully automated segmentation. There was no significant difference between segmentation methods for the T_2 relaxation times for femoral cartilage ($P = 0.641$), tibial cartilage ($P = 0.512$), or patellar cartilage ($P = 0.407$).

The average time for training the SegNet network and the average computing time for cartilage segmentation for each individual image data set are listed in Table 3. The overall training time required until training converges was on the order of days for all image data sets in the current study because of our hardware setup. However, cartilage segmentation was efficient with most of the computation time spent performing 3D simplex deformable modeling. The total cartilage segmentation time using the CNN network was on the order of a few minutes.

DISCUSSION

Our newly developed fully automated segmentation method integrates a deep CED network and 3D simplex deformable modeling approach to improve the efficiency of musculoskeletal tissue segmentation while maintaining low computational costs and short segmentation times. Our segmentation results in the SKI10 challenge suggest that the combined deep CED

and deformable method can achieve high segmentation accuracy with performance superior to most currently used state-of-theart segmentation methods. The deep learning-based method is also time efficiency with average segmentation time on the order of minutes. This is advantageous over atlas-based (15) and model-based (12, 13) segmentation methods because they usually require more computing time (i.e., in the range of several hours) caused by highly computational and intensive registration and feature matching steps (10). In addition, atlas-based and model-based methods require an assumption of mostly normal joint anatomy, which is often not the case in subjects with advanced osteoarthritis. Therefore, in subjects with severe anatomical abnormalities as a result of joint degeneration, accurate bone and cartilage segmentation using these methods can be challenging (10). For deep learning methods, the only limitation is the diversity of training data sets. Given sufficient training data sets including both normal subjects and subjects with varying degrees of joint degeneration, the deep learning-based method should perform better than atlas-based and model-based methods, especially on subjects with severe abnormalities of bone and cartilage.

Not only is our method suitable to the tissue contrast of the fat-suppressed T_1-weighted SPGR sequence used in the SKI10 challenge, it also is applicable to different image contrasts and different spatial resolutions of both morphologic and quantitative MR techniques. Moreover, SegNet method has shown to be more favorable in segmenting bone and cartilage compared against another CED approach (i.e., U-Net) in these data sets. The new segmentation method could potentially be used in clinical practice and osteoarthritis research studies to provide rapid and accurate morphologic and quantitative assessment of the articular cartilage within the knee joint.

Our selection of CNN network differs from other studies and is suitable for performing rapid and accurate cartilage and bone segmentation on musculoskeletal MR images. SegNet is a deep network structure with a total of 32 convolutional layers, much deeper than other applications (22–25,28), which allows the method to learn more abstract and higher-level image features and thereby understand complex image contrasts and varying spatial relationships. Networks with deep structure and small convolutional filters are favorable for tissue segmentation. This is likely attributed to the fact that deeper structure allows more non-linear transformation of the data and is less prone to over-fitting with small convolutional filters (22,26,32). In addition, SegNet performs fast multi-class tissue labeling for an entire image within a single processing pass. Therefore, there is no need for multi-stage training (29) and other supporting aids for region proposals (46) and patch extraction (22,26). Training on whole image slices also has the advantage that allows integrating within the network regularization methods that process large image neighborhoods. Such an example is the integration of a fully connected CRF as an recurrent neural network model on top of a CNN, with the two trained together end-to-end with regular back-propagation algorithm (47). In patch-based methods, this coupled training is less intuitive, as the CRF would potentially need to regularize neighborhoods of voxels larger than a small patch and can, instead, be applied only as a post-processing step (26,48).

The SegNet network uses 2D convolutional filters that limit image features into a single 2D plane. This is likely suboptimal for segmenting 3D object in MR images. The currently

applied regularization in 3D is implemented via the deformable model. As shown in our case for a mesh-based method, the proposed 3D simplex deformable modeling is helpful in maintaining the overall shape of musculoskeletal structures and provides desirable performance in delineating smooth boundaries at tissue interfaces. This step is important for accurately segmenting complex knee joint structures, especially on highly anisotropic images. To further improve the performance, one straight-forward method is to extend SegNet with 3D convolutional filters. However, this approach significantly increases computing load. Another approach is to apply 2D convolutional network into multiple image planes in a similar fashion as described by Prasoon et al (28). Extending SegNet in dimensions perpendicular to the acquired image plane might achieve better performance with appropriate label fusion technique and at isotropic image resolution. An alternative method is to carry out refinement in 3D with a postprocessing step such as 3D CRF (26).

Finally, in comparison with other CED networks, the SegNet is shown to outperform U-Net in segmenting bone and cartilage. This higher performance by SegNet is possibly attributed to the up-sampling via the use of stored indices from max-pooling instead of making the network to learn up-sampling from the scratch, like UNet does. However, this comparison cannot be extrapolated to other studies using U-Net because originally proposed data augmentation was disabled for U-Net in our study to provide a fair comparison. Indeed, U-Net performed considerably well in comparison to other state-of-the-art cartilage segmentation methods. Nevertheless, the CED structure including SegNet and U-Net with a paired encoder and decoder demonstrated an overall good performance in segmenting musculoskeletal tissues.

Training the CNN network requires a long period of time. Our results show that although the segmentation phase is time-efficient, the training phase requires a considerable amount of time (i.e., on the order of days) for a typical training data set using a regular desktop computer (Table 3). From the computing perspective, this could potentially be alleviated by investigating fast parallel computing devices such as by using a high-end graphics processing unit card, graphics processing unit cluster, or distributed system. From the algorithmic perspective, alternative approaches for improving training efficiency in future studies are to use transfer learning from non-medical image to medical image applications (49) and fine-tuning network parameters from pre-trained networks instead of training from scratch using new image data sets (50). Another benefit of transfer learning and fine-tuning is the potential advantage of allowing training on small data sets (51,52). More complex but effective networks contain huge numbers of free parameters to train and therefore require significantly larger numbers of annotated medical images (52). Acquisition of large-scale medical image data sets is typically costly, and manual annotation requires a considerable amount of effort. The applicability of transfer learning between different MR image tissue contrasts and spatial resolutions, between different anatomical structures, and even between different image modalities, are challenging questions that need to be addressed. A theoretically optimal protocol for transfer learning that improves training efficiency while maintaining high segmentation performance has yet to be developed and requires further investigation.

Our study has several limitations. Our study only carried out training and testing of the new fully automated segmentation method on the cartilage and bone of the knee joint. Although our initial results are promising, further research is needed to validate our method for segmenting other anatomical structures within the knee joint and within other joints of the body that are of interest in clinical practice and osteoarthritis research studies. Although our method should be readily adaptable for segmenting any anatomic structure, some structures such as the articular cartilage of the hip joint could pose a challenge because of the thin cartilage and closely opposing articular surfaces. Another limitation is that our study only compared SegNet with one CED network U-Net. Comparison with other networks is challenging because of the fact that the network configuration and objectives are different and highly specific for each individual application. Finally, our feasibility study did not involve transfer learning and fine-tuning of the CNN network, and the network parameters were trained from randomly initialized values. Further investigations are needed to optimize CNN network structures and to investigate image acquisition scheme with optimal image contrast as training data sets for segmentation applications in musculoskeletal imaging.

CONCLUSIONS

In conclusion, our study has developed and evaluated a new fully automated musculoskeletal tissue segmentation method that integrates deep CNN and 3D simplex deformable approaches to improve the accuracy and efficiency of cartilage and bone segmentation within the knee joint. Our segmentation method was shown to provide rapid and accurate cartilage and bone segmentation in time periods short enough for routine use in clinical practice and osteoarthritis research studies. Additional efforts are currently underway to improve the segmentation algorithm by using 3D CED network and fully connected CRF and to reduce training time by applying transfer learning and fine-tuning. Further study is also planned to evaluate this segmentation method on multiple anatomical structures. These techniques may further improve the efficiency of the segmentation method while maintaining high accuracy and repeatability and low computational costs.

REFERENCES

1. Centers for Disease Control and Prevention. Arthritis prevalence and activity limitations--United States, 1990. JAMA 1992;272:346–347.

2. Felson DT. An update on the pathogenesis and epidemiology of osteoarthritis. Radiol Clin North Am 2004;42:1–9. [PubMed: 15049520]

3. Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis Rheum 1995;38:1134–1141. [PubMed: 7639811]

4. Quatman CE, Hettrich CM, Schmitt LC, Spindler KP. The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review. Am J Sports Med 2011;39:1557–1568. [PubMed: 21730207]

5. Conaghan PG, Felson D, Gold G, Lohmander S, Totterman S, Altman R. MRI and non-cartilaginous structures in knee osteoarthritis. Osteo-arthritis Cartilage 2006;14 Suppl A:A87–94.

6. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed 2006;19: 822–854. [PubMed: 17075958]

7. Eckstein F, Cicuttini F, Raynauld JP, Waterton JC, Peterfy C. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. Osteoarthritis Cartilage 2006;14 Suppl A:A46–75.

8. McWalter EJ, Wirth W, Siebert M, von Eisenhart-Rothe RM, Hudelmaier M, Wilson DR, Eckstein F. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthritis Cartilage 2005;13:48–53. [PubMed: 15639637]

9. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116–1128. [PubMed: 16545965]

10. Pedoia V, Majumdar S, Link TM. Segmentation of joint and musculo-skeletal tissue in the study of arthritis. MAGMA 2016;29:207–221. [PubMed: 26915082]

11. Fripp J, Crozier S, Warfield SK, Ourselin S. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. IEEE Trans Med Imaging 2010;29:55–64. [PubMed: 19520633]

12. Seim H, Kainmueller D, Lamecker H, Bindernagel M, Malinowski J, Zachow S. Model-based auto-segmentation of knee bones and cartilage in mri data. Medical Image Analysis for the Clinic: a Grand Challenge. In Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2010), Beijing, China, 2010 p 215–223.

13. Vincent G, Wolstenholme C, Scott I, Bowes M. Fully automatic segmentation of the knee joint using active appearance models. Medical Image Analysis for the Clinic: a Grand Challenge. In Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2010), Beijing, China, 2010 p 224–230.

14. Tamez-Pena JG, Farber J, Gonzalez PC, Schreyer E, Schneider E, Totterman S. Unsupervised segmentation and quantification of anatomical knee features: data from the osteoarthritis initiative. IEEE Trans Biomed Eng 2012;59:1177–1186. [PubMed: 22318477]

15. Dam EB, Lillholm M, Marques J, Nielsen M. Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. J Med Imaging (Bellingham) 2015;2:024001. [PubMed: 26158096]

16. Piplani MA, Disler DG, McCauley TR, Holmes TJ, Cousins JP. Articular cartilage volume in the knee: semiautomated determination from three-dimensional reformations of MR images. Radiology 1996;198:855–859. [PubMed: 8628883]

17. Gougoutas AJ, Wheaton AJ, Borthakur A, Shapiro EM, Kneeland JB, Udupa JK, Reddy R. Cartilage volume quantification via Live Wire segmentation. Acad Radiol 2004;11:1389–1395. [PubMed: 15596377]

18. Kshirsagar AA, Watson PJ, Tyler JA, Hall LD. Measurement of localized cartilage volume and thickness of human knee joints by computer analysis of three-dimensional magnetic resonance images. Invest Radiol 1998;33:289–299. [PubMed: 9609488]

19. Stammberger T, Eckstein F, Michaelis M, Englmeier KH, Reiser M. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. Magn Reson Imaging 1999;17:1033–1042. [PubMed: 10463654]

20. Heimann T, Morrison B, Styner M, Niethammer M, Warfield S. Segmentation of knee images: a grand challenge. Medical Image Analysis for the Clinic: a Grand Challenge. In Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2010), Beijing, China, 2010 p 207–214.

21. Greenspan HG, Bram van, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 2016;35:1153–1159.

22. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in mri images. IEEE Trans Med Imaging 2016;35:1240–1251. [PubMed: 26960222]

23. Brosch T, Tang L, Yoo Y, Li D, Traboulsee A, Tam R. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans Med Imaging 2016;35:1229–1239. [PubMed: 26886978]

24. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Isgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging 2016;35:1252–1261. [PubMed: 27046893]

25. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med Image Anal 2016;30:108–119. [PubMed: 26917105]

26. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61–78. [PubMed: 27865153]

27. Antony J, McGuinness K, Connor NEO, Moran K. Quantifying radio-graphic knee osteoarthritis severity using deep convolutional neural networks. arXiv preprint arXive:1609.02469.

28. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Med Image Comput Comput Assist Interv 2013;16(Pt 2):246–253. [PubMed: 24579147]

29. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. arXiv preprint arXiv: 1605.06211.

30. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. arXiv preprint arXiv 2015: 1505.04597.

31. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561.

32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

33. Delingette H General object reconstruction based on simplex meshes. Int J Comput Vision 1999;32:111–146.

34. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. arXiv preprints arXive 2014:14082014.

35. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

36. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010 p 807–814.

37. Lorensen WE, Cline HE. Marching cubes: a high resolution 3D surface construction algorithm. SIGGRAPH Comput Graph 1987;21:163–169.

38. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 1990;12:629–639.

39. Young IT, van Vliet LJ. Recursive implementation of the Gaussian filter. Signal Process 1995;44:139–151.

40. Cohen LD, Cohen I. A finite element method applied to new active contour models and 3D reconstruction from cross sections. In Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan, 1990 p 587–591.

41. Zarins ZA, Bolbos RI, Pialat JB, Link TM, Li X, Souza RB, Majumdar S. Cartilage and meniscus assessment using T1rho and T2 measurements in healthy subjects and patients with osteoarthritis. Osteoarthritis Cartilage 2010;18:1408–1416. [PubMed: 20696262]

42. Stehling C, Baum T, Mueller-Hoecker C, Liebl H, Carballido-Gamio J, Joseph GB, Majumdar S, Link TM. A novel fast knee cartilage segmentation technique for T2 measurements at MR imaging--data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2011;19:984–989. [PubMed: 21515391]

43. Jarret K, Kavukcuoglu K, Ranzato M, LeCun Y. What is the best multistage architecture for object recognition? In Proceedings of the 12th IEEE International Conference on Computer Vision, Kyoto, Japan, 2009: 2146–2153.

44. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. arXiv preprint arXiv:1502.01852.

45. Bottou L. Large-scale machine learning with stochastic gradient descent; 19th International Conference on Computational Statistics; Heidelberg, Germany. 2010. 177–186.

46. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. arXiv preprint arXiv:1505.04366.

47. Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z, Du D, Huang C, Torr PHS. Conditional random fields as recurrent neural networks. In Proceedings of the 2015 International Conference on Computer Vision, Santiago, Chile, 2015 p 1529–1537.

48. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:1412.7062.

49. Shin HC, Le L, Kim L, Seff A, Yao J, Summers RM. Interleaved text/image deep mining on a large-scale radiology database. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 2015 p 1090–1099.

50. Girshick R, Donahue J, Darrell T, Malik J. Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans Pattern Anal Mach Intell 2016;38:142–158. [PubMed: 26656583]

51. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. arXiv preprint arXiv:1403.6382.

52. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. arXiv preprint arXiv:1602.03409.

53. Wang Z, Donoghue C, Rueckert D. Patch-based segmentation without registration: application to knee MRI In: Wu G, Zhang D, Shen D, Yan P, Suzuki K, Wang F, editors. Machine learning in medical imaging: 4th international workshop, MLMI 2013. Cham: Springer International Publishing; 2013 p 98–105.

54. Lee S, Shim H, Park SH, Yun ID, Lee SU. Learning local shape and appearance for segmentation of knee cartilage in 3D MRI. Medical Image Analysis for the Clinic: a Grand Challenge. In Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2010), Beijing, China, 2010 p 231–240.

55. Yin Y, Williams R, Anderson DD, Sonka M. Hierarchical decision framework with a priori shape models for knee joint cartilage segmentation. Medical Image Analysis for the Clinic: a Grand Challenge. In Proceedings of the 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2010), Beijing, China, 2010 p 241–250.
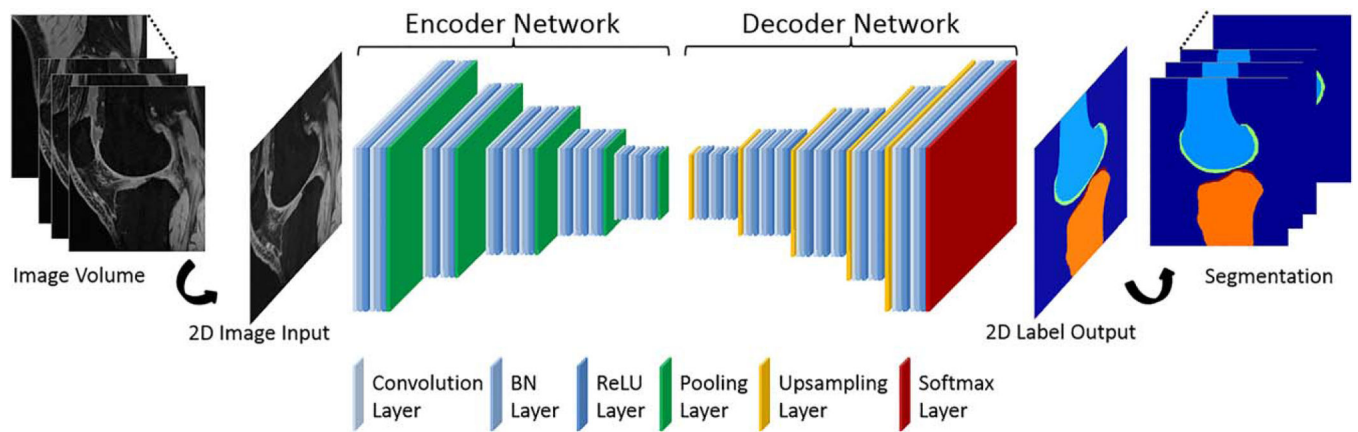
FIG. 1.
An illustration of the SegNet CNN architecture. SegNet contains an encoder network and a decoder network. The final output of this network is high resolution pixel-wise tissue classification.
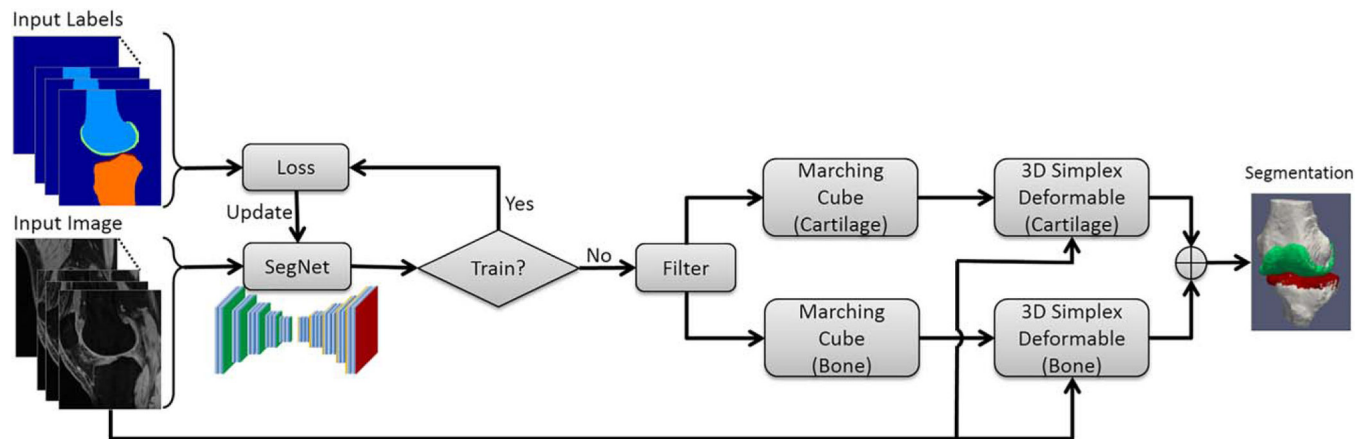
FIG. 2.
Flowchart of our proposed fully automated segmentation method. The complete process contains a training phase and a testing phase. SegNet is first trained with training images and labels. The well-trained SegNet and 3D simplex deformable modeling are then applied in the testing phase to carry out fully automated segmentation for the test images. In the testing phase, the 3D pixel-wise tissue labels from the SegNet output are passed to iterative processing filter to fill holes and remove small isolated objects using a connected-component filter. The processed labels are then sent to 3D simplex deformable process for each individual objects, and the final 3D segmentation is generated by combining all the deformed objects.
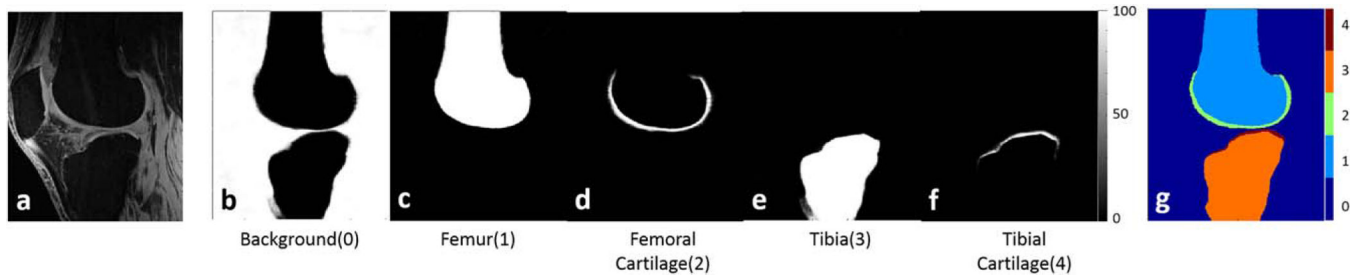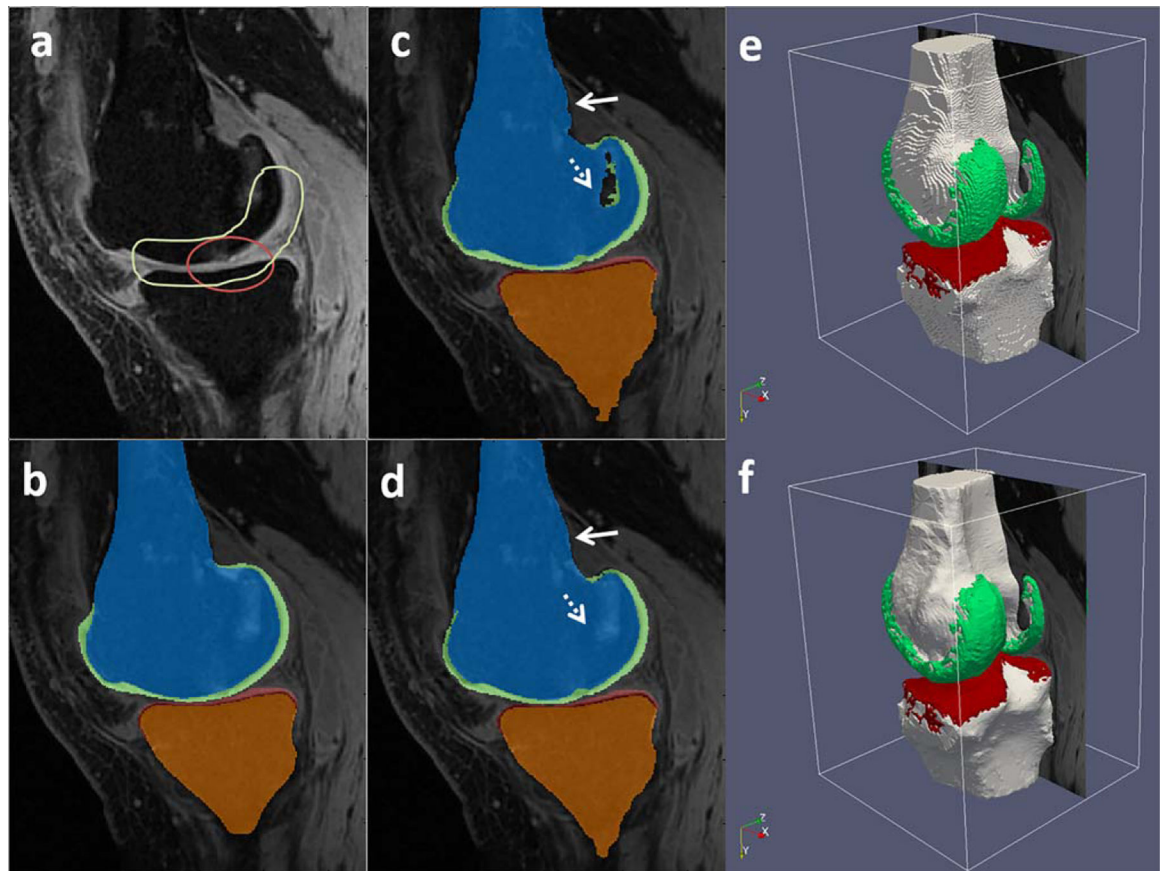
FIG. 3.
Example of the pixel-wise class possibility maps obtained from soft-max layer in SegNet.
(a) A sagittal slice fat-suppressed T_1-weighted SPGR image for one subject in the SKI10
image data set. (b–f) Pixel-wise probability maps (%) for background, femur, femoral
cartilage, tibia, and tibial cartilage, respectively. (g) Pixel-wise classification map where
each individual pixel has the class index with highest class probability.
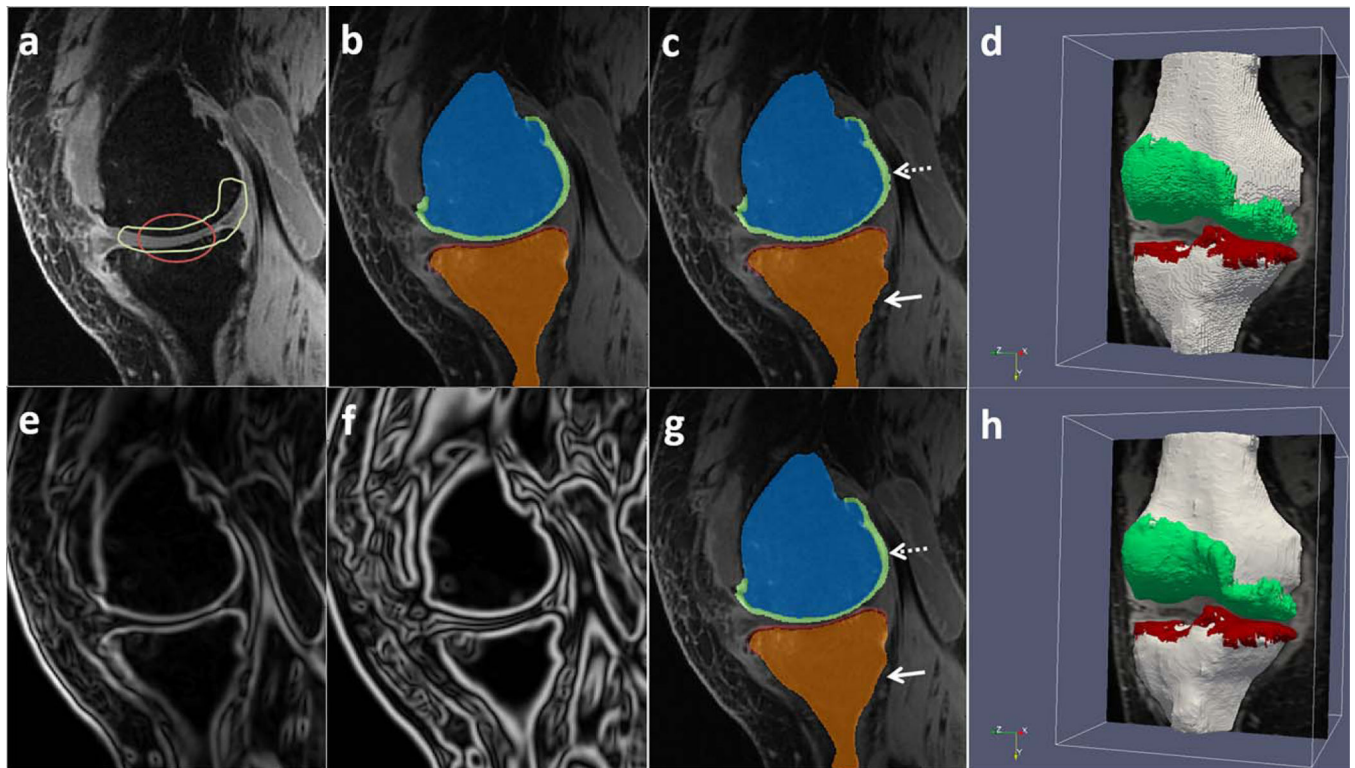
FIG. 4.
Example of the segmentation performance for one subject in the SKI10 image data set with severe cartilage loss. (**a**) A sagittal slice fat-suppressed T_1-weighted SPGR image with ROIs for femoral cartilage (green) and tibial cartilage (red). (**b**) Reference mask of the manual drawn segmentation carried out by experts from the SKI10 challenge.(**c**) Segmentation result directly from SegNet. (**d**) 3D rendering of the segmentation directly result from SegNet. (**e** and **f**) The intermediate filtered images obtained from Perona-Malik anisotropic diffusion filter and recursive Gaussian filter in 3D simplex deformable modeling process. (**g**) Final segmentation result obtained from our proposed method. (**h**) 3D rendering of the final segmentation result.

FIG. 5.

Example of the segmentation performance for one subject in the SKI10 image data set with a bone edema lesion. (**a**) A sagittal slice fat-suppressed T_1-weighted SPGR image with ROIs for femoral cartilage (green) and tibial cartilage (red). (**b**) Reference mask of the manual drawn segmentation carried out by experts from the SKI10 challenge. (**c**) Segmentation result directly from SegNet. (**d**) Final segmentation result obtained from our proposed method. (**e**) 3D rendering of the segmentation result directly from SegNet. (**f**) 3D rendering of the final segmentation result.
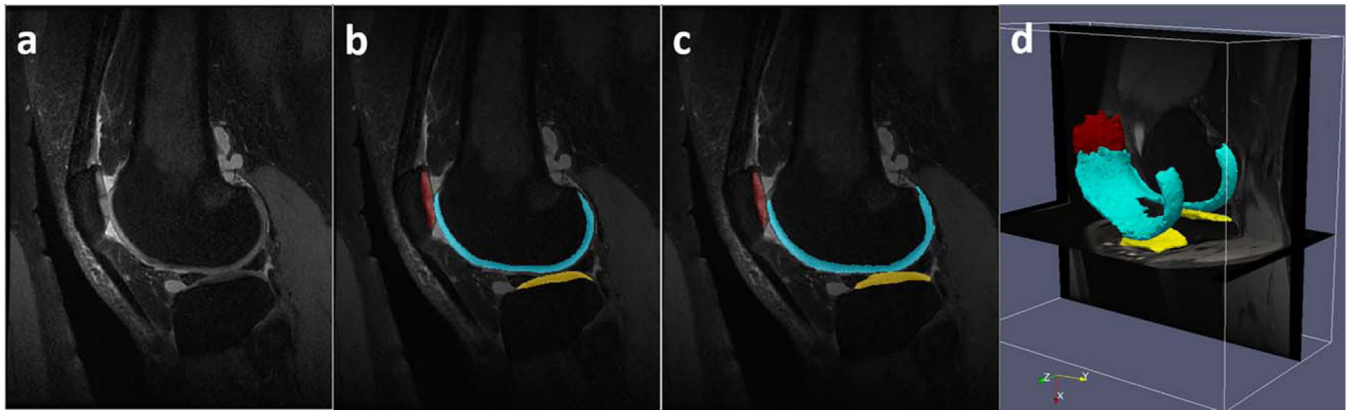
FIG. 6.
Example of the segmentation performance for one subject in the fat-suppressed 3D-FSE image data set. (**a**) A sagittal slice fat-suppressed 3D-FSE image. (**b**) Reference mask of the manual drawn segmentation. (**c**) Final segmentation result obtained from our proposed method with SegNet. (**d**) 3D rendering of the final segmentation result.
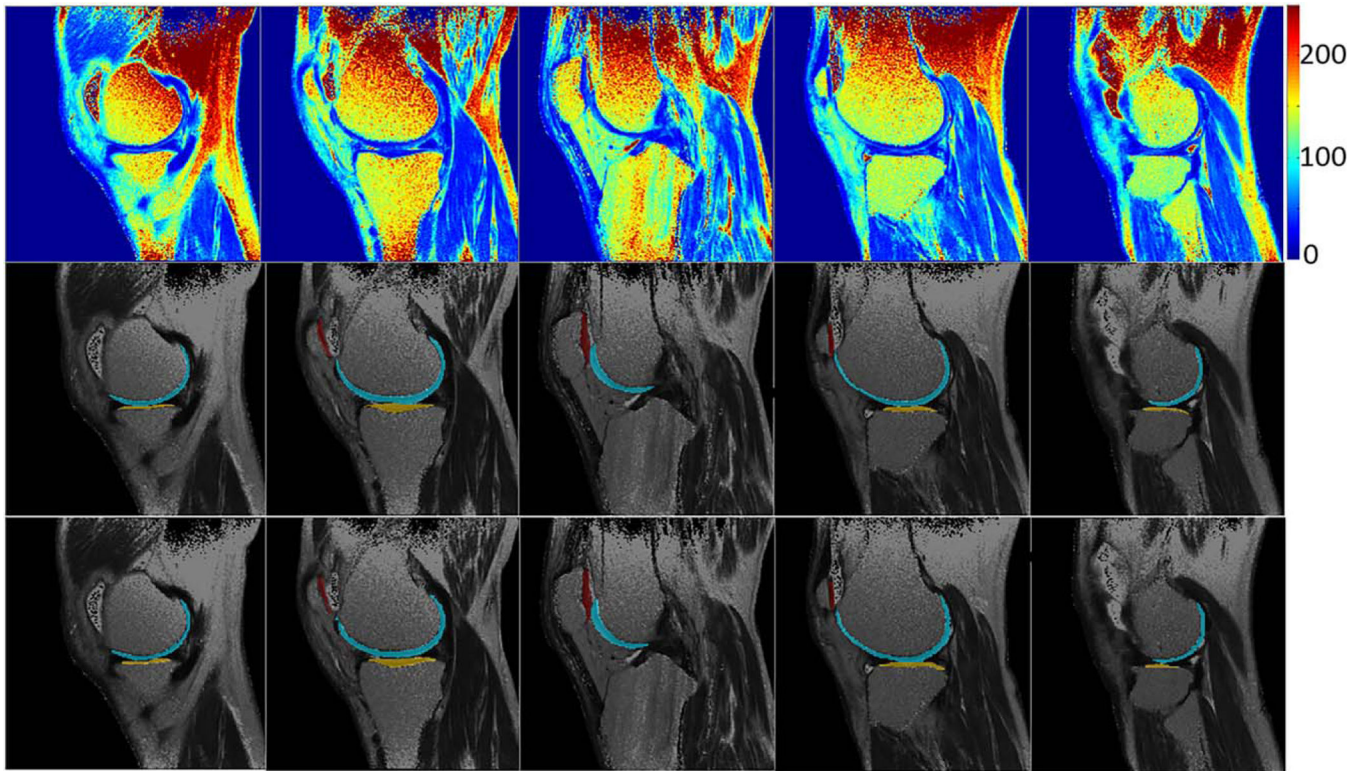
FIG. 7.
Example of the segmentation performance for one subject in the T_2 mapping image data set. Top: multiple sagittal slices T_2 maps generated using the multiple-echo single-exponential signal fitting. Middle: reference mask of the manual drawn segmentation using the DST method. Bottom: final segmentation result obtained from our proposed method with SegNet.

Table 1

Comparison of Segmentation Accuracy (Average Value±SD) of Our Proposed Method and Multiple State-of-the-Art Methods in the SKI10 Competition.

Team (Reference)	Total Scr	Femur Bone		Tibia Bone		Femur Cartilage		Tibia Cartilage	
		AvgD (mm)	RMSD (mm)	AvgD (mm)	RMSD (mm)	VOE (%)	VD (%)	VOE (%)	VD (%)
Biomediq (15)	67.1 ± 8.0	0.68 ± 0.22	1.25 ± 0.41	0.50 ± 0.18	0.91 ± 0.35	26.9 ± 6.0	0.8 ± 13.5	25.1 ± 6.7	0.41 ± 13.4
SegNet + deformable	64.1 ± 9.5	0.56 ± 0.12	1.08 ± 0.21	0.50 ± 0.14	1.09 ± 0.28	28.4 ± 6.9	8.1 ± 12.3	33.1 ± 7.1	–1.2 ± 17.4
SegNet	60.9 ± 10.5	0.70 ± 0.34	1.89 ± 1.25	0.51 ± 0.29	1.48 ± 1.16	29.1 ± 6.4	4.6 ± 11.5	33.4 ± 9.3	1.6 ± 18.2
BioMedIA (53)	56.5 ± 9.2	0.92 ± 0.52	1.78 ± 1.07	0.91 ± 0.92	1.87 ± 1.60	33.4 ± 6.6	0.0 ± 13.2	34.3 ± 6.4	–1.8 ± 16.2
ZIB (12)	54.4 ± 8.8	1.02 ± 0.22	1.54 ± 0.30	0.84 ± 0.19	1.24 ± 0.28	34.0 ± 12.7	7.7 ± 19.2	29.2 ± 8.6	–2.7 ± 18.2
U-Net + deformable	53.0 ± 8.9	0.79 ± 0.12	1.24 ± 0.17	0.80 ± 0.24	1.38 ± 0.47	34.0 ± 9.6	3.6 ± 15.9	34.4 ± 9.2	6.0 ± 19.6
Imorphics (13)	52.3 ± 8.6	0.88 ± 0.24	1.49 ± 0.44	0.74 ± 0.21	1.21 ± 0.34	36.3 ± 5.3	–25.2 ± 10.1	34.6 ± 7.9	–9.5 ± 18.8
SNU_SPL (54)	51.9 ± 10.0	1.37 ± 0.50	2.56 ± 1.06	0.77 ± 0.38	1.40 ± 0.84	30.6 ± 5.4	10.0 ± 14.8	34.0 ± 7.9	2.3 ± 17.6
U-Net	51.1 ± 12.7	1.16 ± 0.55	2.01 ± 1.21	0.94 ± 0.63	2.20 ± 1.40	35.1 ± 5.2	5.5 ± 18.7	34.2 ± 9.3	5.5 ± 19.5
UliibiKnee (55)	34.1 ± 7.8	1.69 ± 0.35	2.97 ± 0.58	2.57 ± 0.53	4.36 ± 0.88	31.2 ± 9.1	–7.3 ± 19.3	33.5 ± 11.0	3.6 ± 19.7

AvgD, average symmetric surface distance; RMSD, root mean square symmetric surface distance; SD, standard deviation; VD, volumetric difference; VOE, volumetric overlap error.

Our proposed method with SegNet achieved high cartilage and bone segmentation accuracy and ranked the second top overall performance based on the SKI10 scoring metrics.

Table 2

Segmentation Accuracy (Average Value ± SD) for the Processed Data Sets.

Data set	Team	Femoral Cartilage		Tibial Cartilage		Patellar Cartilage	
		VOE (%)	VD (%)	VOE (%)	VD (%)	VOE (%)	VD (%)
SKI10	SegNet	29.1 ± 6.4	4.6 ± 11.5	33.4 ± 9.3	1.6 ± 18.2	n/a	n/a
	SegNet + deformable	28.4 ± 6.9	8.1 ± 12.3	33.1 ± 7.1	−1.2 ± 17.4	n/a	n/a
	U-Net	35.1 ± 5.2	5.5 ± 18.7	34.2 ± 9.3	5.5 ± 19.5	n/a	n/a
	U-Net + deformable	34.0 ± 9.6	3.6 ± 15.9	34.4 ± 9.2	6.0 ± 19.6	n/a	n/a
3D-FSE	SegNet	34.6 ± 6.3	11.5 ± 13.7	38.2 ± 5.1	−2.3 ± 17.2	21.9 ± 6.5	5.9 ± 12.7
	SegNet + deformable	33.1 ± 5.2	11.2 ± 10.4	35.9 ± 8.1	6.1 ± 14.2	22.0 ± 7.5	−3.5 ± 14.3
	U-Net	37.2 ± 7.4	13.6 ± 15.9	40.2 ± 8.3	6.1 ± 15.8	24.9 ± 5.5	8.7 ± 13.4
	U-Net + deformable	34.4 ± 6.9	11.4 ± 12.7	38.3 ± 7.9	2.1 ± 16.7	22.8 ± 7.4	5.7 ± 11.4
T$_2$ Map	SegNet	27.5 ± 7.3	7.9 ± 16.2	33.1 ± 9.2	−2.8 ± 14.8	26.1 ± 5.7	5.7 ± 9.5
	SegNet + deformable	24.7 ± 5.8	4.6 ± 13.9	28.9 ± 7.1	3.5 ± 16.4	24.2 ± 4.3	5.7 ± 11.6
	U-Net	32.5 ± 6.2	−6.5 ± 14.7	35.1 ± 10.3	7.8 ± 19.7	27.6 ± 6.4	7.1 ± 9.4
	U-Net + deformable	29.1 ± 7.2	7.5 ± 16.9	32.7 ± 9.3	6.8 ± 18.4	25.6 ± 7.3	6.9 ± 10.6

SD, standard deviation; VD, volumetric difference; VOE, volumetric overlap error.

Table 3

Training and Segmentation Time Using SegNet and Deformable Modeling for Multiple Image Data Sets.

Data Set	Training (h)	Segmentation	
		SegNet (min)	Deformable (min)
SKI10	48.3	0.98	3.2
3D-FSE	30.9	0.31	2.0
T_2 Map	14.7	0.11	0.14