



Published in final edited form as:

Magn Reson Med. 2018 December ; 80(6): 2759–2770. doi:10.1002/mrm.27229.

Deep convolutional neural network for segmentation of knee joint anatomy

Zhaoye Zhou¹, Gengyan Zhao², Richard Kijowski², and Fang Liu²

¹Department of Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota

²Departments of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin

Abstract

Purpose: To describe and evaluate a new segmentation method using deep convolutional neural network (CNN), 3D fully connected conditional random field (CRF), and 3D simplex deformable modeling to improve the efficiency and accuracy of knee joint tissue segmentation.

Methods: A segmentation pipeline was built by combining a semantic segmentation CNN, 3D fully connected CRF, and 3D simplex deformable modeling. A convolutional encoder-decoder network was designed as the core of the segmentation method to perform high resolution pixel-wise multi-class tissue classification for 12 different joint structures. The 3D fully connected CRF was applied to regularize contextual relationship among voxels within the same tissue class and between different classes. The 3D simplex deformable modeling refined the output from 3D CRF to preserve the overall shape and maintain a desirable smooth surface for joint structures. The method was evaluated on 3D fast spin-echo (3D-FSE) MR image data sets. Quantitative morphological metrics were used to evaluate the accuracy and robustness of the method in comparison to the ground truth data.

Results: The proposed segmentation method provided good performance for segmenting all knee joint structures. There were 4 tissue types with high mean Dice coefficient above 0.9 including the femur, tibia, muscle, and other non-specified tissues. There were 7 tissue types with mean Dice coefficient between 0.8 and 0.9 including the femoral cartilage, tibial cartilage, patella, patellar cartilage, meniscus, quadriceps and patellar tendon, and infrapatellar fat pad. There was 1 tissue type with mean Dice coefficient between 0.7 and 0.8 for joint effusion and Baker's cyst. Most musculoskeletal tissues had a mean value of average symmetric surface distance below 1mm.

Conclusion: The combined CNN, 3D fully connected CRF, and 3D deformable modeling approach was well-suited for performing rapid and accurate comprehensive tissue segmentation of the knee joint. The deep learning-based segmentation method has promising potential applications in musculoskeletal imaging.

Keywords

conditional random field; deep learning; deformable model; image segmentation; knee; musculoskeletal imaging

1 | INTRODUCTION

Osteoarthritis (OA) is one of the most prevalent chronic diseases worldwide with the knee being most commonly affected joint.^{1,2} OA research has traditionally focused on cartilage. However, significant confusion exists in explaining the relationship between cartilage degeneration and perceived pain^{3,4} and differences in rates of cartilage loss in the human population.^{5,6} Furthermore, the mechanisms of pain and mechanical dysfunction in OA are not completely understood but are believed to involve multiple interrelated pathways involving all joint structures.^{7–9} Therefore, OA is now considered to be a “whole-organ” disease and not a disease isolated to cartilage or any other joint structure.¹⁰

Magnetic resonance imaging has been used in population-based studies over the past decade to provide important information regarding structural features associated with joint pain and the incidence and progression of OA.¹¹ Various scoring systems, such as the whole-organ magnetic resonance score (WORMS),¹² the Boston Leeds osteoarthritis knee score (BLOKS),¹³ and the MR osteoarthritis knee score (MOAKS),¹⁴ have been developed to provide a semiquantitative assessment of the severity of each structural feature of knee joint degeneration. Although semi-quantitative methods can be used to evaluate all musculoskeletal tissues including cartilage, bone, synovium, meniscus, and tendon, the scoring systems are time-consuming, subjective, and highly dependent on the level of reader expertise.

Quantitative measures of joint degeneration have also been used in OA research studies and have the advantages of being objective and highly reproducible with a greater dynamic range for assessing tissue degeneration than semiquantitative grading scales.¹⁵ Quantitative MR imaging was first used in OA research studies to measure cartilage thickness^{16–18} and relaxation characteristics reflecting cartilage composition and microstructure.^{19–21} More recently, quantitative assessment of bone shape,^{22–24} bone marrow edema lesion size,^{25–27} synovial fluid volume,²⁸ meniscus shape and position,^{29–31} infrapatellar fat pad volume and signal intensity,^{32–34} and muscle bulk^{35–37} have provided important information regarding the association between these structural features and joint pain and the incidence and progression of OA.

Segmentation of musculoskeletal tissues is the crucial first step in the processing pipeline to acquire quantitative measures of joint degeneration from MR images. Traditionally, tissue segmentation has been performed manually in which a user delineates the boundaries of each joint structure on each MR image slice that is extremely time-consuming with its efficiency and repeatability influenced by the level of user expertise.³⁸ Fully automated tissue segmentation techniques based on active shape modeling³⁹ and atlas databases⁴⁰ have been recently developed and have shown to have promising results for segmenting cartilage and bone. However, these methods rely on a priori knowledge of knee shapes from many

healthy subjects and require high computation costs because of the need to access pre-stored databases and continuously compare current detected knee shapes to pre-stored shapes.

In recent years, adaptation of deep learning techniques, particularly convolutional neural networks (CNN), have been used for musculoskeletal tissue segmentation from MR images. Prasoon et al.⁴¹ proposed a multi-planar approach to segment cartilage in which 2D CNN networks were performed in 3 orthogonal image planes with the final segmentation achieved by combining predictions from each network. The technique was found to outperform a 3D method for segmenting tibial cartilage in patients with knee OA. More recently, Liu et al.⁴² introduced an approach combining highly efficient convolutional encoder-decoder (CED) networks with image intensity post-processing for semantic multiple class tissue segmentation. The technique was found to be superior to state-of-the-arts model-based and atlas-based methods for segmenting cartilage and bone in patients with knee OA.

CNN approaches have shown promising results for segmenting cartilage and bone. However, the adaptation of CNN methods for rapid and accurate segmentation of all joint structures that may be sources of pain in patients with OA has yet to be investigated. Inspired by the work from Liu et al.,⁴² we propose a method combining an improved CED network, 3D fully connected conditional random field (CRF), and 3D simplex deformable modeling to perform comprehensive tissue segmentation of the knee joint. The purpose of this study is to describe and validate the method for segmenting cartilage, bone, tendon, meniscus, muscle, infrapatellar fat pad, and joint effusion and Baker's cyst.

2 | METHOD

2.1 | Convolutional encoder-decoder architecture

The key component in the automated segmentation method is a deep CED network. The CED network is an adaptation of a network structure used in studies by Liu et al.⁴² for segmenting cartilage and bone in musculoskeletal MR images and for labeling skull in PET/MR attenuation correction.⁴³ Figure 1 shows a schematic demonstration of the deep learning network. The CED network is featured by a connected encoder network and a decoder network. The encoder network acts as both a compressor and feature-detector of the input image data set. Specifically, the popular VGG16 network⁴⁴ is used as the encoder because it has been proven to be an effective network of feature extraction in CED-based medical image applications.^{42,43} Each unit layer of the VGG16 encoder consists of a convolutional layer with a varying set of 2D filters, a batch normalization (BN),⁴⁵ a rectified-linear unit (ReLU) activation,⁴⁶ and a max-pooling process for reducing the data dimensions. This unit layer is repeated 5 times in the encoder network to achieve sufficient data compression.

To output pixel-wise labels, a decoder network is applied following the encoder network. Because the decoder network is the reverse process of the encoder, its structure is mirrored from the encoder network and consists of convolutional layers that shares similar structure as the encoder network. But in the decoder network, the upsampling layer takes the place of the max-pooling layer in the encoder network to consecutively upsample the image features and increase output feature resolution. The last layer of the decoder network is a multi-class

softmax classifier that generates class probabilities for each individual pixel with the exact same resolution as the input images.

In addition, a symmetric shortcut connection (SC) between the encoder and decoder network is added to improve the CED network by promoting the labeling performance. The role of the shortcut connection is to preserve sufficient image details during the max-pooling process in the encoder and to enhance the training efficiency in the deep network structure.⁴⁷ A total of 4 symmetric shortcut connections are generated based on the full pre-activation scheme in the deep residual network configuration between the encoder layers and decoder layers.⁴⁷ Figure 1 illustrates the detailed structure of the proposed CED network and the shortcut connection.

2.2 | Fully connected 3D conditional random field

The CED network trains and predicts image data in a slice-by-slice fashion that results in a stack of 2D class probability maps. Although a 2D CNN is highly efficient in processing high in-plane resolution images, the inter-slice contextual information might not be fully handled. Therefore, irregular labels such as holes and small isolated objects are likely to be generated in regions with ambiguous image contrast. To effectively assign the labels to voxels with similar image intensity values and to take into account the 3D contextual relationships among voxels, a fully connected 3D conditional random field (CRF) is applied to fine-tune the segmentation results from the CED networks.^{48,49}

In the fully connected 3D CRF process, a maximum a posteriori (MAP) inference is defined over the 3D volume of the whole knee joint. In the model, the probability results for each label from the CED network are used to generate the unary potential on each voxel, and the original 3D knee image volume is used to calculate the pairwise potentials on all pairs of voxels. The iterative CRF optimization is carried out by minimizing the Gibbs energy defined as

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j), \quad (1)$$

where x_i and x_j are the labels assigned to the i^{th} and j^{th} voxel, respectively. The values of i and j range from 1 to the total number of voxels and $\psi_u(x_i)$ is the unary potential defined as the negative logarithm of the probability for a particular label from the CED soft-max prediction. The pairwise potential $\psi_p(x_i, x_j)$ is defined as

$$\begin{aligned} \psi_p(x_i, x_j) = \mu(x_i, x_j) & \left[\omega_1 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right) \right. \\ & \left. + \omega_2 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right) \right], \end{aligned} \quad (2)$$

where p_i and p_j are the voxel locations and I_i and I_j are the image intensity values. The pairwise potential $\Psi_p(x_i, x_j)$ is a function of the appearance kernel and the smoothness kernel that are expressed as 2 exponential terms in the Equation 2. The appearance kernel, which is the first exponential term in the equation, assumes voxels close to each other or having similar image intensity values tend to share the same label. The extent of each effect is controlled by θ_α and θ_β . The smoothness kernel, the second exponential term, removes isolated small regions, and its effect on the pairwise potential is controlled by θ_γ . The weights for the appearance kernel and the smoothness kernel are determined by ω_1 and ω_2 , respectively. The compatibility function, $\mu(x_i, x_j)$ is set as the Potts model,

$$\mu(x_i, x_j) = \begin{cases} 1 & \text{if } x_i \neq x_j \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In this study, a highly efficient algorithm proposed by Krähenbühl et al.⁴⁹ is applied to make the complex inference with a tremendous number of pairwise potentials practical in 3D image volume. As a result, the inference algorithm for 3D fully connected CRF is linear to the number of variables and sublinear to the number of edges in the model, thereby increasing the computing efficiency of the whole knee joint refinement process.

2.3 | 3D deformable model for bone and cartilage

Based on the results from the fully connected 3D CRF process, each voxel is assigned with a label with the highest class probability. Voxels sharing the same class index are defined as being the same tissue. The class labels of the entire 3D knee joint volume are used as the initial tissue classification labels. In particular, a smooth and well-defined boundary is desirable for cartilage and bone. Therefore, 3D deformable modeling is implemented for cartilage and bone refinement. A marching cube algorithm is applied to extract the boundary of each individual segmentation objects for cartilage and bone.⁵⁰ Simplex meshes are used in the model to represent the 3D shape that allows for robust and efficient 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 that is expressed as⁵¹

$$m \frac{d^2 V}{dt^2} = -\lambda \frac{dV}{dt} + F_{in} + F_{ex}, \quad (4)$$

where m is the vertex mass, λ is the damping factor ranging from 0 to 1 that 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 the mesh and image boundaries. Similar to the processing described in Liu et al.,⁴² the numerical solution of the Equation 4 can be obtained using the central finite differences with an explicit scheme at a discrete time point (t) as described in⁵¹

$$V^{t+1} = V^t + (1 - \lambda)(V^t - V^{t-1}) + \alpha F_{in} + \beta F_{ex}, \quad (5)$$

where α and β are tunable internal and external force factors, respectively. Empirical experiment results have shown that a stable iterative deformation can be achieved when α has a value below 0.5. The β is typically selected to be a value smaller than 1.⁵¹ Deformable refinement is individually performed for the femoral, tibial, and patellar cartilage and the bone. The deformed results are combined for final tissue segmentation in a 3D class label volume.

2.4 | Full segmentation pipeline

The schematic diagram of the segmentation approach is illustrated in Figure 2. The 3D image volume of the knee joint is dissembled into a stack of 2D image slices. In the training phase, the 2D image slices are used as the input of the CED network and are compared to the corresponding pixel-wise class labels in the training data. The training loss of the CED network is determined by multi-class cross-entropy loss weighted by the inverse of class occurring frequency (i.e., larger weight for structure with smaller volume).⁵² The weights of the CED network keep getting updated and recorded until the number of iterations reaches a predefined maximum step. Thereafter, the weights of the network used during the testing phase are selected from the iteration wherein the training loss is the lowest among all. In the testing phase, the well-trained CED network is used as a frontend segmentation classifier to segment the testing 2D images and to generate tissue class probabilities for each pixel. The fully connected 3D CRF model is further applied to refine the overall segmentation results by improving the label assignment for voxels with similar contrast and taking into account the 3D contextual relationships. In particular, the processed labels representing cartilage and bone are discretized into 3D simplex meshes by using the marching cube algorithm and are sent to the 3D simplex deformable process. In the deformable process, each individual segmentation object is refined to preserve smooth tissue boundary and maintain overall anatomical geometry. The final 3D segmentation is obtained by merging all the segmentation objects.

The segmentation algorithm is implemented in a hybrid programming environment. The CED is designed and coded using the Keras package with Tensorflow as the computing backend.⁵³ The image processing filters, CRF models, and 3D simplex deformable subroutine are implemented using MATLAB (version 2013a, MathWorks, Natick, MA) and libraries from the Insight Segmentation and Registration Toolkit (ITK) and the Visualization Toolkit (VTK) from KitWare Inc. (Clifton Park, NY).

2.5 | Image data sets

The study was performed in compliance with HIPAA regulations, with approval from our Institutional Review Board and with all subjects signing informed consent. A sagittal frequency selective fat-suppressed 3D fast spin-echo (3D-FSE) sequence was performed on the knee joint of 20 subjects with knee OA (12 males and 8 females with an average age of 58 years) using a 3T scanner (Discovery MR750, GE Healthcare, Waukesha, WI) and 8-

channel phased-array extremity coil (InVivo, Orlando, FL). The imaging parameters included a 2216 ms TR, 23.6 ms TE, 90° flip angle, ± 75 kHz bandwidth, 16cm FOV, 384×384 matrix size, 0.416×0.416mm in-plane resolution, 2mm slice thickness, and 45 image slices.

Manual tissue segmentation was carried out by a musculoskeletal research assistant using the conventional segmentation feature in MATLAB (version 2013a) under the supervision of a fellowship-trained musculoskeletal radiologist with 15 yrs of clinical experience. A multi-class mask with 13 classes was created for each 3D-FSE image slice of each subject with the following values: 0=background, 1=femur, 2=femoral cartilage, 3=tibia, 4=tibial cartilage, 5=patella, 6=patellar cartilage, 7=meniscus, 8=quadriceps and patellar tendons, 9=muscle, 10=synovial fluid-filled joint effusion and Baker's cyst, 11=infrapatellar fat pad, and 12=other non-specified tissues.

2.6 | Network training and post-processing

All training and evaluation were carried out on a personal desktop computer hosting a 64-bit Ubuntu Linux operating system. Computing hardware included an Intel Xeon W3520 quad-core CPU, 32GB DDR3RAM, and one Nvidia GeForce GTX 1080 Ti graphic card with total 3584 cores, and 11GB GDDR5RAM.

All the input 2D images were first cropped to enclose as much of the knee joint as possible while removing excessive image background. The images were then resampled to 320×224 matrix size using bilinear interpolation before they were sent to the CED network for training and evaluation. Image augmentation was used to compensate for the small number of training data sets by creating 3 repeats of the training data using 2D image translation, shearing and rotation provided by the standard ImageDataGenerator function in Keras package.⁵⁴ Transfer learning was also implemented using 60 sagittal 3D T₁-weighted spoiled gradient recalled-echo (3D-SPGR) knee image data sets from The Segmentation of Knee Images 2010 (SKI10, <http://www.ski10.org>) competition, one of the featured image segmentation challenges hosted by the MICCAI conference in 2010, which included pixel-wise cartilage and bone labels annotated by imaging experts.⁵⁵ The CED network was first trained on the SKI10 image data sets using a randomly initialized network from a scheme described by He et al.⁵⁶ Training was then performed on the 3D-FSE image data sets using the pre-trained network from the SKI10 data sets. The CED network was optimized using Adam algorithm⁵⁷ with a fixed learning rate of 0.001 and trained in a mini-batch manner with 10 image slices in a single mini-batch. A total iteration steps corresponding to 20 and 50 epochs for the SKI10 and 3D-FSE data sets, respectively, were carried out for training convergence. In addition, a leave-one-out cross-validation was performed on the 3D-FSE image data sets. Twenty training folds were performed in total. In each training fold, 19 subjects with a total of 3420 2D image slices including 855 original image slices and 2565 augmented image slices were used for training and 1 remaining subject was used for evaluation. The evaluation subject was altered in each training fold to ensure every subject was used once as evaluation subject after all 20 training folds.

The parameters for 3D fully connected CRF process were empirically selected and included $\theta_\alpha=5$, $\theta_\beta=5$, $\theta_\gamma=3$, the weight factor $\omega_1=3$, and $\omega_2=3$ in the Equation 2. A total 10 iteration

steps were used in the inference of the 3D fully connected CRF process. The parameters for the 3D deformable process were selected from Liu et al.⁴² and included an internal force factor $\alpha=0.3$, external force factor $\beta=0.01$, and damping factor $\lambda=0.65$ in Equation 5. A total of 50 iteration steps were used to ensure surface smoothness and to maintain overall object shape.

To compare the performance with networks using 3D convolutional filters, 2 state-of-the-art 3D CNNs, deepMedic⁵⁸ and V-Net,⁵⁹ were adapted in current study for simultaneously segmenting all knee joint tissue structures. The deepMedic is a 3D patch-based method using multiscale image processing pathways and the V-Net is a full 3D volume-based method. Except the number of classes, both networks were configured at default network settings proposed in the original papers. The deepMedic was trained for 35 epochs (20 sub-epochs per epoch with 1000 samples in each sub-epoch) at the same sampling rate for all tissue classes. Because of the extremely high GPU memory demand for the V-Net, the input full 3D image volume was resampled into 128×128×64 matrix size as suggested by the original V-Net paper, and the network was trained for a total of 500 epochs with a batch size of 2. The same pre-train using SKI10 data sets and the leave-one-out cross-validation were also performed for deepMedic and V-Net to provide an unbiased comparison to the proposed method.

2.7 | Evaluation of segmentation accuracy

To evaluate the accuracy of tissue segmentation, the Dice coefficient was used for each individual joint structure and was defined as

$$DC = \frac{2|S \cap R|}{|S| + |R|}, \quad (6)$$

where S and R represent the segmentation by the CNN approach and the manual reference segmentation ground truth, respectively. The Dice coefficient ranged between 0 and 1 with a value of 1 indicating a perfect segmentation and a value of 0 indicating no overlap at all. The volumetric overlap error (VOE) and volumetric difference (VD) were also calculated to evaluate the accuracy of cartilage segmentation. The VOE was defined as

$$VOE = 100 \left(1 - \frac{|S \cap R|}{|S \cup R|} \right), \quad (7)$$

with smaller VOE value indicating a more accurate segmentation. The VD was defined as

$$VD = 100 \frac{||S| - |R||}{|R|}, \quad (8)$$

to indicate the size difference of the segmented cartilage. The VOE and VD values were calculated within a ROI that was drawn in each of 3 consecutive central slices on the medial and lateral tibial plateau, medial and lateral femoral condyles, and patella on the 3D-FSE

image data sets. In addition, the average symmetric surface distance (ASSD) was also calculated for evaluating the surface overlap between the segmented mask and the ground truth for each individual joint structure. The ASSD was defined as

$$\text{ASSD} = \frac{\sum_{s \in \partial(S)} \min_{r \in \partial(R)} \|s - r\| + \sum_{r \in \partial(R)} \min_{s \in \partial(S)} \|r - s\|}{|\partial(S)| + |\partial(R)|}, \quad (9)$$

where $\partial(\cdot)$ means the boundary of the segmentation set.

3 | RESULTS

The overall training time required until training converged was on the order of hours for the image data sets given the computing hardware in the current study. The training loss curves are shown in Supporting Information Figure S1. More specifically, the total training time was ~1.5h for each fold in the leave-one-out cross-validation. However, tissue segmentation was rather fast with a mean computing time of 0.2 min, 0.8 min, and 3 min for the CED, the 3D fully connected CRF, and the 3D deformable modeling process, respectively, for each image volume.

Figure 3 shows the bar plot of the Dice coefficient and ASSD values for each individual segmented joint structure in the 20 subjects with knee OA. All musculoskeletal tissues after the full process had a mean Dice coefficient above 0.7. There were 4 tissue types with high Dice coefficients above 0.9 including the femur (mean±SD: 0.970±0.010), tibia (0.962±0.015), muscle (0.932±0.024), and other non-specified tissues (0.913±0.017). There were 7 tissue types with Dice coefficients between 0.8 and 0.9 including the femoral cartilage (0.806±0.062), tibial cartilage (0.801±0.052), patella (0.898±0.033), patellar cartilage (0.807±0.101), meniscus (0.831±0.031), quadriceps and patellar tendon (0.815±0.029), and infrapatellar fat pad (0.882±0.040). There was 1 tissue type with a Dice coefficient between 0.7 and 0.8 for joint effusion and Baker's cyst (0.736±0.069). The averaged segmentation error and SD of VOE for the proposed method was 31.7%±6.1%, 32.4%±4.3% and 20.8%±5.9% for femoral cartilage, tibial cartilage, and patellar cartilage, respectively. The average segmentation error and SD of VD for the proposed method was 13.7%±8.4%, -5.6%±9.3%, and 4.8%±11.4% for femoral cartilage, tibial cartilage, and patellar cartilage, respectively. Most musculoskeletal tissues after the full process had a mean ASSD below 1mm. There were 4 tissue types with low ASSD below 0.5mm including the femur (0.332±0.503mm), femoral cartilage (0.499±0.332mm), tibia (0.377±0.412mm), and tibial cartilage (0.452±0.138mm). There were 5 tissue types with ASSD between 0.5mm and 1mm including the patella (0.570±0.273mm), patellar cartilage (0.663±0.297mm), meniscus (0.565±0.217mm), tendon (0.794±0.277mm), and other non-specified tissues (0.548±0.194mm). There were 3 tissue types with an ASSD above 1mm including muscle (1.067±1.127mm), joint effusion and Baker's cyst (3.534±2.692mm), and infrapatellar fat pad (1.100±1.178mm). Both deepMedic and V-Net resulted in less accurate segmentation results in comparison with the proposed method on the basis of Dice coefficient and ASSD. Within these 2 3D CNN methods, full volume-based V-Net performed significantly better

than patch-based deepMedic and provided close results to the proposed method for segmenting all knee joint tissue structures.

Figure 4 shows sagittal examples of tissue segmentation performed on the 3D-FSE images of the knee joint in 2 subjects using the CED network only, the CED network combined with 3D fully connected CRF, and the CED network combined with both CRF and 3D deformable modeling. The full figure including axial and coronal views is shown in Supporting Information Figure S2. The first subject of Figure 4 is a 56-year-old male with mild knee OA. The segmentation results from the CED network demonstrated good agreement with the overall contours of the ground truth. However, there were mislabeled tissue voxels in a portion of the joint effusion, meniscus and femur (white arrows) that was corrected using 3D fully connected CRF. The use of 3D deformable modeling resulted in further smoothing of the tissue boundaries of cartilage and bone. There was good agreement between the final segmentation and the overall shape of the ground truth for all joint structures. The second subject of Figure 4 is a 64-year-old male with severe knee OA. Segmentation of this image data set was challenging because of the aliasing artifact and metallic artifact from prior knee surgery that degraded image contrast for the femur, femoral cartilage, and patellar tendon. The segmentation results from the CED network demonstrated good agreement with the overall contours of the ground truth for most joint structures. However, there were mislabeled tissue voxels in a portion of the femur, femoral cartilage, infrapatellar fat pad, patellar tendon, and meniscus (white arrows) that was corrected using 3D fully connected CRF. The use of 3D deformable modeling resulted in smoothing of the tissue boundaries of cartilage and bone. There was good agreement between the final segmentation and the overall shape of the ground truth for all joint structures despite the presence of image artifacts and advanced tissue degeneration.

Figure 5 shows examples of 3D rendered models for all joint structures in a 46-year-old female with mild knee OA created using the segmented tissue masks. The 3D rendered models demonstrated good estimation of the complex anatomy for both regular shaped structures such as bone, cartilage, meniscus, and tendon, and irregular shaped structures such as joint effusion and Baker's cyst.

4 | DISCUSSION

Our study demonstrated the feasibility of using a deep learning-based approach for efficient and accurate segmentation of all joint structures in patients with knee OA. Our method incorporated a deep CED network combined with pixel-wise label refinement using 3D fully connected CRF and contour-based 3D deformable refinement using 3D simplex modeling. Liu et al.⁴² first demonstrated the ability of a CED network combined with 3D simplex modeling to efficiently and accurately segment cartilage and bone within the knee joint. We modified the previously described approach to include 3D fully connected CRF that is a post-processing algorithm that uses the consistency of image intensity for correction of mislabeled tissue voxels. The segmentation accuracy of our deep CED network, quantified using VOE and VD measurements for femoral, tibial, and patellar cartilage, was comparable to the segmentation accuracy for cartilage reported by Liu et al.⁴² However, our study also showed that the deep learning-based approach could achieve high segmentation efficiency

and accuracy not only for cartilage and bone but also for other joints structures that may be sources of pain in patients with knee OA including tendon, meniscus, muscle, infrapatellar fat pad, and joint effusion and Baker's cyst.

Our selection of a CED network was critical for performing multi-class segmentation of musculoskeletal tissues from MR images. The proposed CED network was a deep network structure that allowed the model to learn more complex image contrasts and features. Deeper networks are preferable for multi-class tissue segmentation at a large number of classes because they are better suited to perform complex transformation and are less sensitive to over-fitting of data. In our study, we evaluated the network using 13 classes including 12 different tissue types and one background class. Knee joint structures with large volumes such as muscle and with regular geometrical shapes such as the femur and tibia achieved the highest segmentation accuracy. Smaller structure such as the meniscus and more complex and irregular shaped structures such as joint effusion and Baker's cyst demonstrated lower segmentation accuracy. However, larger training data sets could be used in the future to improve the segmentation accuracy for these smaller and more complex and irregular shaped knee joint structures.

Our deep learning-based approach required the use of post-processing algorithms for accurate musculoskeletal tissue segmentation. The CED network used 2D convolutional filters that limit the features within an image slice and cause segmentation bias when contrast is inconsistent across all slices in an image data set. Multi-planar CNN, usually referred to as 2.5D method, which performs CNN prediction on multiple orthogonal image planes can be helpful for incorporating 3D spatial information.⁴¹ This strategy performs quite well at isotropic or nearly isotropic image resolution but is suboptimal for anisotropic image data sets where slice thickness of image is much larger than in-plane voxel size. Implementation of 3D convolution in CNN is also suggested by many studies,^{58–60} however, 3D CNN methods typically require extensive computing resources such as abundant GPU memory and therefore are limited to high-end computing setup. For full image size multi-class segmentation, as described herein, downsampling input image might be necessary to make full volume-based 3D CNN feasible, which can also degrade segmentation performance as shown in current study for the V-Net. Another requirement for favorable training of full volume-based 3D CNN is larger training data set because each 3D volume is now treated as a single training sample. The 3D patch-based CNN method requires less training data set and GPU memory. Despite this method might perform quite well in segmenting small number of classes in volumetric data (e.g., brain lesion segmentation),⁵⁸ the limited spatial context in each patch and relative shallow network structure prevent such method from accurately segmenting complex tissue structures with a large number of classes as demonstrated in the deepMedic results. In comparison, 3D fully connected CRF was implemented in our deep learning-based approach as a post-processing step for correction of mislabeled tissue voxels by the 2D CNN. The improvement in tissue segmentation was similar to the results of previous studies in which 3D CRF was used to regularize segmentation boundaries at tissue interfaces.^{58,61,62} 3D simplex deformable modeling was also used as an additional post-processing step to further smooth the boundaries of cartilage and bone.

One major advantage of deep learning-based approaches for tissue segmentation is that they do not require prior knowledge of structural shape. Traditional model-based and atlas-based segmentation approaches estimate the boundaries of joint structures by computing spatial structural differences to a normalized shape reference. These methods require an assumption of mostly normal joint anatomy and face challenges for image data sets with substantial structural variability and local feature differences because of artifact-induced image degradation and advanced tissue degeneration. In contrast, deep learning-based approaches can provide accurate segmentation even in the presence of image artifacts and advanced tissue degeneration and are suitable for segmenting structures without normalized shape references such as joint effusion and Baker's cyst. In addition, deep learning-based approaches for tissue segmentation are highly time efficient. Despite a relative long training time that needs to be performed only once, the segmentation is highly efficient with an average processing time on the order of minute for all knee joint structures. This is advantageous in comparison to traditional atlas-based and model-based segmentation approaches that usually involve 1 or multiple time-consuming registration steps for each individual joint structure. Therefore, performing segmentation of multiple structures using these methods is extremely time-consuming and impractical for use in large population-based OA research studies.

The results of our study serve as a first step to provide quantitative MR measures of musculoskeletal tissue degeneration for OA research studies. The deep learning-based segmentation method could be used to create 3D rendered models for all joint structures that may be sources of pain in patients with knee OA. These models could be directly used to measure bone shape,^{22–24} synovial fluid volume,²⁸ meniscus shape and position,^{29–31} infrapatellar fat pad volume and signal intensity,^{32–34} and muscle bulk.^{35–37} Further postprocessing of the 3D rendered bone models with an image intensity thresholding technique could be used to isolate and measure the volume of bone marrow edema lesions.^{25–27} The 3D rendered models could also be used as masks to superimpose over quantitative MR parameter maps to assess the composition and microstructure of cartilage,^{19–21} meniscus,⁶³ and tendon.⁶⁴

Our study has several limitations. Our feasibility study only evaluated the accuracy of the CED network for segmenting knee joint structures using 3D-FSE image data sets. Although the 3D-FSE images provided excellent tissue contrast for evaluating all joint structures, further extending the proposed method to segment images with different tissue contrasts is warranted. Training the CED network is computationally expensive and requires a large amount of pixelwise annotated training data sets for each new tissue contrast evaluated. Our study took advantages of the SKI10 image data set and transferred learned features from the SKI10 images to the 3D-FSE images. Although our result using transfer learning agrees well with previous studies where transfer learning was shown to greatly reduce the amount of data necessary for successfully training segmentation network for medical image,⁶⁵ there are still needs to investigate the optimal number of training data that provides the most cost-effective performance in musculoskeletal image segmentation. Additional work is needed to implement transfer learning and network fine-tuning to allow the CED network to accurately segment musculoskeletal tissues on images with different tissue contrasts and in joints other than the knee using smaller training data sets. Another limitation of our study was that it did

not directly compare the CED network with other atlas-based and model-based approaches for musculoskeletal tissue segmentation. Nevertheless, the Dice coefficients and other quantitative segmentation accuracy metrics reported in our study could be used in future studies to compare our CED network with other methods used to segment knee joint structures.

5 | CONCLUSIONS

Our study has described and evaluated a new approach using a deep CED network combined with 3D fully connected CRF and 3D simplex modeling for performing efficient and accurate multi-class musculoskeletal tissue segmentation from MR images. The deep learning-based segmentation method could be used to create 3D rendered models of all joint structures including cartilage, bone, tendon, meniscus, muscle, infrapatellar fat pad, and joint effusion and Baker's cyst that may be sources of pain in patients with knee OA. The results of our study serve as a first step to provide quantitative MR measures of musculoskeletal tissue degeneration in a highly time efficient manner that would be practical for use in large population-based OA research studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

- [1]. 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]
- [2]. Felson DT. An update on the pathogenesis and epidemiology of osteoarthritis. *Radiol Clin North Am.* 2004;42:1–9. [PubMed: 15049520]
- [3]. Link TM, Steinbach LS, Ghosh S, et al. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. *Radiology.* 2003;226:373–381. [PubMed: 12563128]
- [4]. Hunter DJ, March L, Sambrook PN. The association of cartilage volume with knee pain. *Osteoarthritis Cartilage.* 2003;11:725–729. [PubMed: 13129691]
- [5]. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL. Longitudinal study of changes in tibial and femoral cartilage in knee osteoarthritis. *Arthritis Rheum.* 2004;50:94–97. [PubMed: 14730604]
- [6]. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, et al. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. *Arthritis Res Ther.* 2005;8:R21. [PubMed: 16507119]
- [7]. Felson DT, Chaisson CE, Hill CL, et al. The association of bone marrow lesions with pain in knee osteoarthritis. *Ann Intern Med.* 2001;134:541–549. [PubMed: 11281736]
- [8]. Hill CL, Gale DG, Chaisson CE, et al. Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. *J Rheumatol.* 2001;28:1330–1337. [PubMed: 11409127]
- [9]. Sowers MF, Hayes C, Jamadar D, et al. Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis. *Osteoarthritis Cartilage.* 2003;11:387–393. [PubMed: 12801478]
- [10]. Peterfy C, Woodworth T, Altman R. Workshop for Consensus on Osteoarthritis Imaging: MRI of the knee. *Osteoarthritis Cartilage.* 2006;14:44–45.

- [11]. Eckstein F, Kwoh CK, Link TM, OAI investigators. Imaging research results from the Osteoarthritis Initiative (OAI): a review and lessons learned 10 years after start of enrolment. *Ann Rheum Dis*. 2014;73:1289–1300. [PubMed: 24728332]
- [12]. Peterfy CG, Guermazi A, Zaim S, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage*. 2004;12:177–190. [PubMed: 14972335]
- [13]. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds osteoarthritis knee score). *Ann Rheum Dis*. 2008;67:206–211. [PubMed: 17472995]
- [14]. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI osteoarthritis knee score). *Osteoarthritis Cartilage*. 2011;19:990–1002. [PubMed: 21645627]
- [15]. Li Q, Amano K, Link TM, Ma CB. Advanced imaging in osteoarthritis. *Sports Health*. 2016;8:418–428. [PubMed: 27510507]
- [16]. Reichenbach S, Yang M, Eckstein F, et al. Does cartilage volume or thickness distinguish knees with and without mild radiographic osteoarthritis? The Framingham study. *Ann Rheum Dis*. 2010;69:143–149. [PubMed: 19193659]
- [17]. Eckstein F, Maschek S, Wirth W, et al. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. *Ann Rheum Dis*. 2009;68:674–679. [PubMed: 18519425]
- [18]. Le Graverand MPH, Buck RJ, Wyman BT, et al. Change in regional cartilage morphology and joint space width in osteoarthritis participants versus healthy controls: a multicentre study using 3.0 Tesla MRI and Lyon-Schuss radiography. *Ann Rheum Dis*. 2010;69:155–162. [PubMed: 19103634]
- [19]. Liebl H, Joseph G, Nevitt MC, et al. Early T2 changes predict onset of radiographic knee osteoarthritis: data from the osteoarthritis initiative. *Ann Rheum Dis*. 2015;74:1353–1359. [PubMed: 24615539]
- [20]. Prasad AP, Nardo L, Schooler J, Joseph GB, Link TM. T1 ρ and T2 relaxation times predict progression of knee osteoarthritis. *Osteoarthritis Cartilage*. 2013;21:69–76. [PubMed: 23059757]
- [21]. Liu F, Choi KW, Samsonov A, et al. Articular cartilage of the human knee joint: in vivo multicomponent T2 analysis at 3.0 T. *Radiology*. 2015;277:477–488. [PubMed: 26024307]
- [22]. Barr AJ, Dube B, Hensor EMA, et al. The relationship between clinical characteristics, radiographic osteoarthritis and 3D bone area: data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2014;22:1703–1709. [PubMed: 25278079]
- [23]. Hunter D, Nevitt M, Lynch J, et al. Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. *Ann Rheum Dis*. 2016;75:1607–1614. [PubMed: 26483253]
- [24]. Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. *Ann Rheum Dis*. 2015;74: 519–525. [PubMed: 24306109]
- [25]. Frobell RB, Roos HP, Roos EM, et al. The acutely ACL injured knee assessed by MRI: are large volume traumatic bone marrow lesions a sign of severe compression injury? *Osteoarthritis Cartilage*. 2008;16:829–836. [PubMed: 18206394]
- [26]. Driban JB, Price L, Lo GH, et al. Evaluation of bone marrow lesion volume as a knee osteoarthritis biomarker - longitudinal relationships with pain and structural changes: data from the Osteoarthritis Initiative. *Arthritis Res Ther*. 2013;15:R112. [PubMed: 24020939]
- [27]. Pang J, Driban JB, Destenaves G, et al. Quantification of bone marrow lesion volume and volume change using semi-automated segmentation: data from the osteoarthritis initiative. *BMC Musculoskelet Disord*. 2013;14:3. [PubMed: 23281825]
- [28]. Gait AD, Hodgson R, Parkes MJ, et al. Synovial volume vs synovial measurements from dynamic contrast enhanced MRI as measures of response in osteoarthritis. *Osteoarthritis Cartilage*. 2016;24:1392–1398. [PubMed: 27038489]

- [29]. Wenger A, Englund M, Wirth W, Hudelmaier M, Kwok K, Eckstein F. Relationship of 3D meniscal morphology and position with knee pain in subjects with knee osteoarthritis: a pilot study. *Eur Radiol*. 2012;22:211–220. [PubMed: 21842432]
- [30]. Roth M, Wirth W, Emmanuel K, Culvenor AG, Eckstein F. The contribution of 3D quantitative meniscal and cartilage measures to variation in normal radiographic joint space width—data from the Osteoarthritis Initiative healthy reference cohort. *Eur J Radiol*. 2017;87:90–98. [PubMed: 28065381]
- [31]. Emmanuel K, Quinn E, Niu J, et al. Quantitative measures of meniscus extrusion predict incident radiographic knee osteoarthritis - data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2016;24:262–269. [PubMed: 26318658]
- [32]. Lu M, Chen Z, Han W, et al. A novel method for assessing signal intensity within infrapatellar fat pad on MR images in patients with knee osteoarthritis. *Osteoarthritis Cartilage*. 2016; 24:1883–1889. [PubMed: 27327783]
- [33]. Steidle-Kloc E, Culvenor AG, Dörrenberg J, et al. Relationship between knee pain and infrapatellar fat pad morphology - a within-and between-person analysis from the Osteoarthritis Initiative. *Arthritis Care Res*. 2018;70:550–557.
- [34]. Ruhdorfer A, Haniel F, Petersohn T, et al. Between-group differences in infra-patellar fat pad size and signal in symptomatic and radiographic progression of knee osteoarthritis vs nonprogressive controls and healthy knees - data from the FNIH Biomarkers Consortium Study and the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2017;25:1114–1121. [PubMed: 28216313]
- [35]. Conroy MB, Kwok CK, Krishnan E, et al. Muscle strength, mass, and quality in older men and women with knee osteoarthritis. *Arthritis Care Res*. 2012;64:15–21.
- [36]. Hart HF, Ackland DC, Pandy MG, Crossley KM. Quadriceps volumes are reduced in people with patellofemoral joint osteoarthritis. *Osteoarthritis Cartilage*. 2012;20:863–868. [PubMed: 22525223]
- [37]. Pan J, Stehling C, Muller-Hocker C, et al. Vastus lateralis/vastus medialis cross-sectional area ratio impacts presence and degree of knee joint abnormalities and cartilage T2 determined with 3T MRI - an analysis from the incidence cohort of the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2011;19:65–73. [PubMed: 21044692]
- [38]. McWalter EJ, Wirth W, Siebert M, et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. *Osteoarthritis Cartilage*. 2005;13: 48–53. [PubMed: 15639637]
- [39]. Frupp 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]
- [40]. 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]
- [41]. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a tri-planar convolutional neural network. *Med Image Comput Comput Assist Interv*. 2013;16(Pt 2):246–253. [PubMed: 24579147]
- [42]. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med*. 2018;79:2379–2391. [PubMed: 28733975]
- [43]. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. *Radiology*. 2018;286:676–684. [PubMed: 28925823]
- [44]. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv e-prints arXiv* 2014: 1409–1556.
- [45]. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv e-prints arXiv* 2015:1502–03167.
- [46]. 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.

- [47]. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. arXiv e-prints arXiv 2016:1603–05027.
- [48]. He X, Zemel RSRS, Carreira-Perpinan MAA, et al. Multiscale conditional random fields for image labeling. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, 2004 p. 695–703.
- [49]. Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials. arXiv e-prints arXiv 2012: 1210–5644.
- [50]. Lorensen WE, Cline HE. Marching cubes: a high resolution 3D surface construction algorithm. Comput Graph (ACM). 1987;21: 163–169.
- [51]. Delingette H General object reconstruction based on simplex meshes. Int J Comput Vis. 1999;32:111–146.
- [52]. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. arXiv e-prints arXiv 2014:1411–4038.
- [53]. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv eprints arXiv 2016:1603–04467.
- [54]. Chollet François. Keras. GitHub 2015 <https://github.com/fchollet/keras>. Published March 27, 2015 Updated February 13, 2018. Accessed March 27, 2015.
- [55]. Heimann T, Morrison B, Styner M, Niethammer M, Warfield S. Segmentation of knee images: a grand challenge. Med Image Comput Comput Assist Interv. 2010:207–214.
- [56]. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. arXiv e-prints arXiv 2015:1502–08152.
- [57]. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv e-prints arXiv 2014:1412–6980.
- [58]. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017; 36:61–78. [PubMed: 27865153]
- [59]. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. arXiv e-prints arXiv 2016:1606–04797.
- [60]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. arXiv e-prints arXiv 2016:1606–06650.
- [61]. Christ PF, Ettlinger F, Grün F, et al. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv e-prints arXiv 2017: 1702–05970
- [62]. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv e-prints arXiv 2014:1412–7062
- [63]. Liu F, Samsonov A, Wilson JJ, Blankenbaker DG, Block WF, Kijowski R. Rapid in vivo multicomponent T2 mapping of human knee menisci. J Magn Reson Imaging. 2015;42:1321–1328. [PubMed: 25847733]
- [64]. Kijowski R, Wilson JJ, Liu F. Bicomponent ultrashort echo time T2* analysis for assessment of patients with patellar tendinopathy. J Magn Reson Imaging. 2017;46:1441–1447. [PubMed: 28263448]
- [65]. van Opbroek A, Ikram MA, Vernooij MW, de Bruijne M. Transfer learning improves supervised image segmentation across imaging protocols. IEEE Trans Med Imaging. 2015;34: 1018–1030. [PubMed: 25376036]

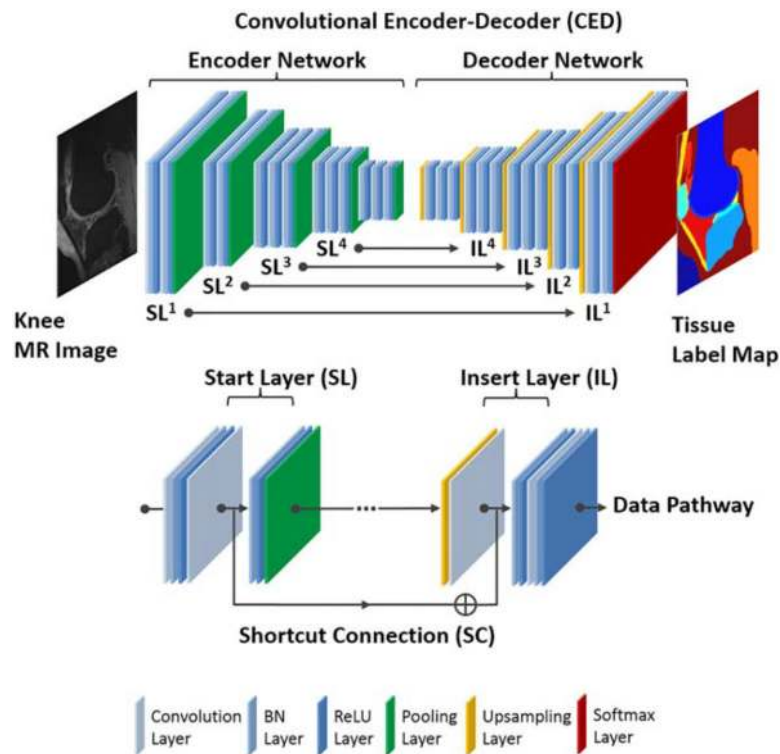
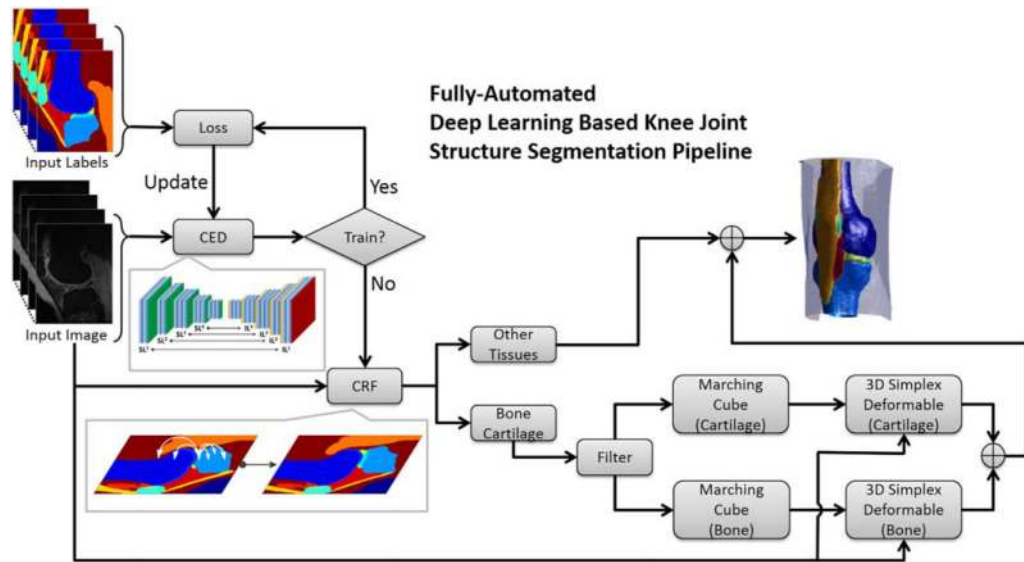
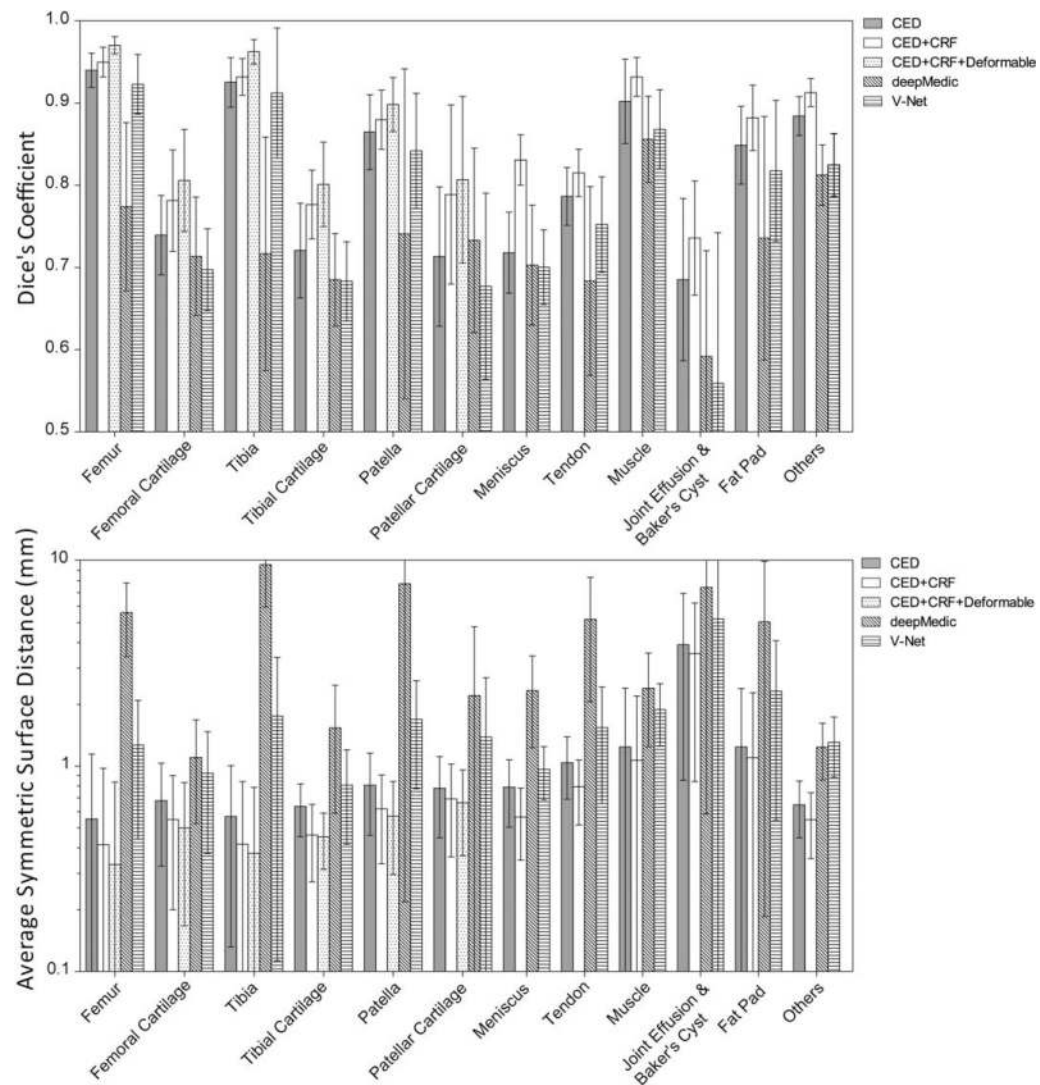
**FIGURE 1.**

Illustration of the CNN architecture, which features a connected encoder network and a decoder network. The encoder network uses typical VGG16 convolutional layers and the decoder uses a mirrored structure of the encoder network with max-pooling replaced by upsampling process. A symmetric shortcut connection (SC) between the encoder and decoder network is added to improve the CED network for promoting the labeling performance. All convolutional filters use 3×3 filter size and the max-pooling layers use a 2×2 window and stride 2. All the input 2D images from the training data set were first cropped to enclose as much of the knee joint as possible while removing excessive image background. The images were then resampled to 320×224 matrix size using bilinear interpolation to match the fixed input size of the CED network

**FIGURE 2.**

Flowchart of the fully automated musculoskeletal tissue segmentation method. The well-trained CED network is used to segment the testing images and to generate tissue class probabilities for each pixel. The CRF model is further applied to promote the label assignment for voxels. The processed labels representing cartilage and bone are discretized by using the marching cube algorithm and are sent to the 3D simplex deformable process. In the deformable process, each individual target is refined to preserve smooth tissue boundary and to maintain overall anatomical geometry

**FIGURE 3.**

Bar plot (mean and SD) of the Dice coefficient and average symmetric surface distance (ASSD) values for each individual segmented joint structure in the 20 subjects for CED, combination of CED and CRF, the combination of CED, CRF and deformable process, deepMedic and V-Net. Note that all musculoskeletal tissues had a mean Dice coefficient above 0.7 by using the full process of proposed method. Most musculoskeletal tissues had a mean value of ASSD below 1mm by using our proposed method

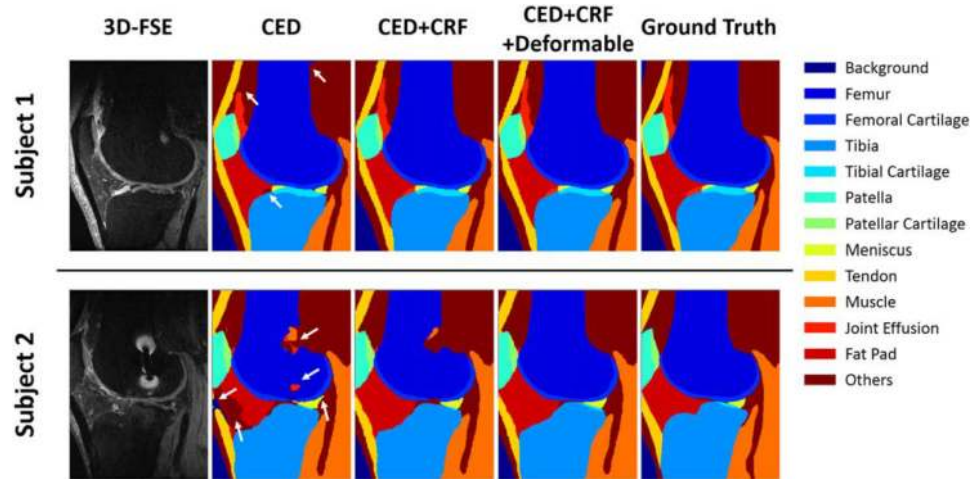


FIGURE 4.

Examples of tissue segmentation performed on the 3D-FSE images in 2 subjects with knee OA using the CED network only, the CED network combined with 3D fully connected CRF, and the CED network combined with both CRF and 3D deformable modeling. Note the improvement in segmentation accuracy achieved by combining the CED network with 3D fully connected CRF and 3D deformable modeling

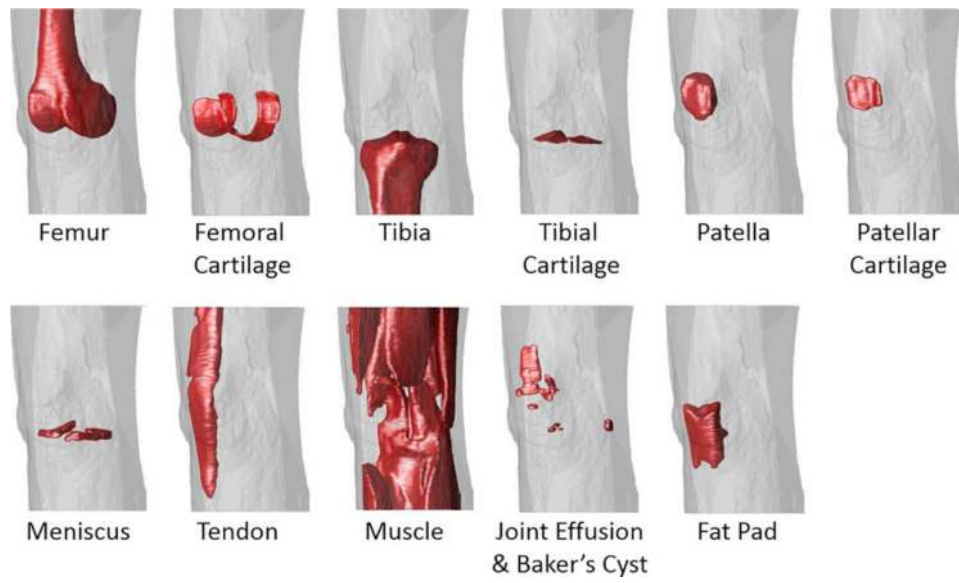


FIGURE 5.
Examples of 3D rendered models for all joint structures in a subject with knee OA created using the segmented tissue masks