Automatic knee cartilage segmentation using fully volumetric convolutional neural networks for evaluation of osteoarthritis

Conference Paper · April 2018

DOI: 10.1109/ISBI.2018.8363705

5 authors, including:

Archit Raj
Indian Institute of Technology Delhi
1 PUBLICATION **1** CITATION
SEE PROFILE

Srikrishnan Viswanathan
GE Healthcare, Bangalore
23 PUBLICATIONS **53** CITATIONS
SEE PROFILE

Bhavya Ajani
Samsung
5 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Karthik Krishnan
Samsung Research Bangalore
26 PUBLICATIONS **211** CITATIONS
SEE PROFILE

AUTOMATIC KNEE CARTILAGE SEGMENTATION USING FULLY VOLUMETRIC CONVOLUTIONAL NEURAL NETWORKS FOR EVALUATION OF OSTEOARTHRITIS

Archit Raj, Srikrishnan Vishwanathan, Bhavya Ajani, Karthik Krishnan and Harsh Agarwal

Samsung R&D Institute India Bangalore Pvt. Ltd., Bangalore, Karnataka, India

ABSTRACT

Automated Cartilage segmentation is essential for improving the performance of advanced Knee Osteoarthritis (OA) assessment due to its convoluted 3D structure. In this paper, we have developed a knee cartilage segmentation algorithm from a high resolution MR volume using a novel 3D- fully Convolutional Neural Network (CNN), called 'µ-Net' coupled with a multi-class loss function. This is, to our knowledge, the first automatic cartilage segmentation method using 3D CNNs. The proposed algorithm performed better than the state-of-the-art algorithm in the MICCAI SKI10 public challenge. We have further applied our proposed algorithm on another similar MR contrast (DESS) provided by *Osteoarthritis Initiative (OAI)* for OA assessment and have presented improved segmentation accuracies. Initial qualitative assessment of segmentation results visually depicts cartilage loss in longitudinal knee MR data.

Index Terms— Knee, MRI, Osteoarthritis, Deep Learning, Segmentation, Cartilage

1. INTRODUCTION

Knee osteoarthritis (OA) is leading cause of global disability [1, 2]. The lifetime risk of developing knee OA is 40% and 47% in men and women, respectively. The risk rises to 60% for those with BMI > 30 [1]. Mechanical wear & tear and, biochemical changes in knee cartilage are the primary cause of OA [3]. Current clinical protocol, detects OA post mechanical damage when it is irreversible. Recent studies have shown that knee OA is reversible if detected at an early stage of biochemical changes in knee cartilage [4]. Therefore, joint analysis of mechanical wear & tear and biochemical changes in knee cartilage is warranted for advancement in the management of knee OA.

Quantitative MRI techniques such as T2 and T1ρ mapping have been shown as potential biomarkers for biochemical changes in cartilage for early detection of OA [5]. Biomarkers for cartilage measurement such as its volume, thickness and surface area are used in clinical practice to assess OA progression [6]. Computations of these biomarkers are sensitive to the accurate cartilage segmentation especially when cartilage loss is tracked over time.

Manual segmentation of cartilage is laborious (average 118/156 min per subject [7]) and subjective because of its thin, elongated and curved structure which spans into several slices. Various methods have been proposed for automatic segmentation of cartilage. These include learning based classifiers using hand crafted features and non-learning based approaches that rely on a statistical distribution of intensity or shape such as Active Shape/Appearance Models, atlas based methods, level sets and graph cuts [8]. In 2010, SKI10 challenge has been held to objectively compare algorithms for knee segmentation on a standardized test bed of 100 MR T1w/T2w volumes with Imorphics being the winner [8, 9] at the time of writing this manuscript. Deep learning based techniques have recently been shown to achieve superior performance in various segmentation tasks [10]. However, due to the memory limitations and long training time tri-planar 2D-CNNs have been used for 3D cartilage segmentation [11]. Recently fully volumetric 3D CNNs with a V-Net architecture and a two class DICE loss function was deployed for segmentation of the prostate [12].

The Osteoarthritis Initiative (OAI) provides multicenter, longitudinal, prospective observational study to promote research in the management of knee osteoarthritis [13]. Annotated 176 DESS MR volumes are also provided for the development of cartilage segmentation algorithms.

In this manuscript we have developed a 3D- fully CNN network; 'µ-Net' because its shape resembles the Greek character, µ. µ-Net employs auxiliary classifiers for deep supervision [14], residual links [15] and fully volumetric CNNs [12] along with a multi-class DICE loss function. Segmentation results from µ-Net were compared on MR T1w volumes from SKI10 challenge [8]. A segmentation result of µ-Net on OAI's MR DESS was also evaluated [13]. Initial qualitative clinical utility assessment was done on OAI dataset.

2. MATERIALS AND METHODS

2.1. Network Architecture: µ- Net

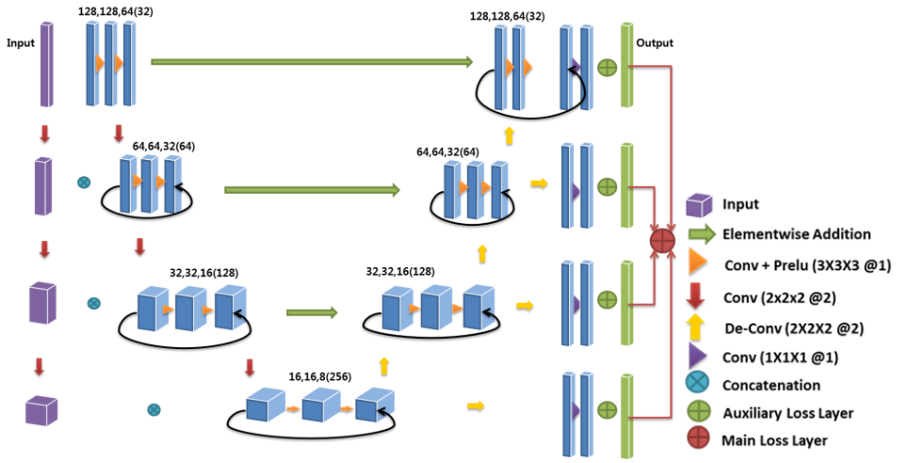

Fig 1: μ- Net used for cartilage segmentation

Fig. 1 shows the network architecture of μ- Net. Inspired by the standard U-Net [16, 17], it has an analysis and a synthesis path each with four resolution steps. In the analysis path, each resolution step contains two 3×3×3 convolutions, followed by a parametric rectified linear unit (PReLU), and a 2×2×2 convolution with strides of two for down-sampling. While in the synthesis path, each resolution step consists of two 3×3×3 convolutions, followed by a PReLU and a deconvolution of 2×2×2 with strides of two for up-sampling.

Short-skip and long-skip connections follow element wise addition operation to carry out the flow of information within the network. Short-skip connections are added within each resolution of the analysis and synthesis path to eliminate the problem of vanishing gradients [15]. Long-skip connections are added across the layers of equal resolution from the analysis to synthesis path enabling high-resolution feature flow in the network [16].

The information lost during down-sampling was avoided by feeding the input in analysis path. It is fed using concatenation operation at each resolution, and the resolution of input is reduced by half using a 2×2×2 convolution with strides of two. Loss layers are added at each step in the synthesis part of the network which acts as an auxiliary classifier, providing a form of deep supervision [14]. All resolution steps in the synthesis path are first up-sampled using deconvolution to the native input resolution, followed by a 1×1×1 convolution that reduces the number of output channels to number of labels and are fed to the auxiliary classifier. The feed from all the auxiliary classifiers is added to a main classifier which controls the learning rate of different auxiliary classifiers as a function of iterations.

Dice loss [11] is modified to accommodate for multiple labels to generate a multi-class (3 for SKI10 and 7 for OAI), N, segmentation. Modified Dice loss, D_k at k^{th} resolution, for auxiliary loss layer (shown as green in Fig 1) is given by,

$$D_k = \frac{2\sum_i^N p_i\, g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}, \frac{\partial D_k}{\partial p_j} = 2\left[\frac{g_j(\sum_i^N p_i^2 + \sum_i^N g_i^2) - 2p_j(\sum_i^N p_i\, g_i)}{(\sum_i^N p_i^2 + \sum_i^N g_i^2)^2}\right]$$

with D_4 being highest resolution. p_i and g_i are probability map and ground truth for i^{th} label, respectively. Loss, D, for main loss layer (shown as red in Fig 1) is given by,

$$D = \sum_{k-1} (\alpha_k * D_k) + \beta D_4$$
$$\alpha_k = \alpha_k \left(1 - \frac{\#(Iteration)}{\#(Total\ Iteration)}\right), \quad \sum_{k-1} \alpha_k + \beta = 1.$$

Weights, α_k are updated with each iteration and weight β for final resolution is designed to make sure that overall dice loss never exceeds 1. In this manuscript, starting weights, α_k and β are initialized to 0.25 in first iteration.

2.2. Data and Labels

SKI10 challenge [8] data was used for the comparison of segmentation from μ-net with the state-of the art technique. It contained 100 3D MR volumes acquired in a multi-site, multi-vendor and multi field-strength setting. T1w or T2w MR volumes were acquired in sagittal acquisitions; with an in-plane resolution of 0.4mm and slice thickness of 1mm. Reference labels of the femoral and tibial cartilage along with respective bones were also provided. Only femoral and tibial cartilage labels were used in this manuscript. Additionally, labels for load bearing regions of tibial and femoral cartilages were provided for evaluation. 80 MR volumes were used for training and 20 were used for testing.

OAI data was used for further validation of the segmentation algorithm to access its use in future clinical studies. OAI data consisted of 176 knee 3D MR volumes (88 patients, one baseline scan and one 12-month follow-up scan) along with reference annotations for 6 structures (femoral cartilage, left and right tibial cartilage, left and right menisci and patellar cartilage). 3D Double Echo Steady State (3D-DESS) MR volumes were acquired in sagittal acquisitions with 160 slices, 0.7mm slice thickness, water excitation for fat suppression, FOV 140×140mm, in-plane acquisition resolution 0.365×0.456mm, TE/TR=4.7/16.3msec, slice partial-Fourier factor 0.75. 140 MR volumes were used for training and 35 were used for testing.

	Dice Score	Avg. HD	VOE		VD	
	Our Method	Our Method	Our Method	Imorphics [9]	Our Method	Imorphics [9]
Femoral Cartilage	0.834 ± 0.011 (0.811,0.856)	0.218 ± 0.023 (0.173, 0.264)	28.302 ± 1.583 (25.198, 31.404)	36.3 ± 5.3	12.504 ± 6.052 (0.641, 24.368)	-25.2 ± 10.1
Tibial Cartilage	0.825 ± 0.010 (0.806, 0.844)	0.226 ± 0.011 (0.204, 0.247)	29.461 ± 1.406 (26.704, 32.216)	34.6 ± 7.9	16.798 ± 8.672 (-0.199, 33.796)	-9.5 ± 18.8

Table 1: Comparison of results (mean ± standard deviation along with 95% CI) on MICCAI datasets with Imorphics [9]

2.3 Training and testing

μ-net is trained independently for the SKI10 and the OAI data. No pre-/post-processing apart from down/up-sampling is done. All the volumes and reference labels are re-sampled to 128×128×64 voxels corresponding to a voxel-spacing of $1\times1\times1.5mm^3$ and fed to the network due to memory limitation of the GPU during training.

Original training dataset was augmented to make the model more robust to new testing datasets. Data augmentation was performed by using random non-linear deformations in space and intensity [12]. Deformation in space was done using a dense deformation field obtained through a 2×2×2 grid of control-points and a B-spline interpolation. Deformation in intensity distribution of the data was done by using histogram matching.

The auxiliary classifiers and lateral up-sampling which were added in the network during training were removed at the time of testing, as these layers and classifiers were used to add additional supervision in the network training for increased accuracy. The output of the last convolutional layer, which has the same resolution as that of input to the network, is passed through a soft-max layer. The Soft-max layer generates a probability map for the background and the N structures (7 for the OAI model and 3 for the SKI10 model) at each voxel in the image. Each of these probability maps are resampled to the original MR image resolution to mitigate the impact of lower resolution images at the input of the network. The voxels are assigned to the class with the maximum probability to generate a multi-label segmentation map.

2.4 Computation time

Training took ~2 days on a Linux machine with Nvidia Tesla K80. During testing, the full pipeline took ~5secs on a, Dell-7910 Xeon 24-2.5GHz with 64GB RAM, Nvidia Tesla K40c, Windows 7 machine while the network execution stage took ~3secs. A modified version of the Caffe library [12] (modified to support 3D fully volumetric CNNs) was used for the entire operation.

3. RESULTS

Evaluation was carried out using 5-fold cross validation (CV). Four metrics, identical to [8] were computed. These are the Dice Score, Volume Overlap Error (VOE), Volume Difference (VD) and Average Hausdorff Distance (Avg. HD),

$$\text{Dice Score} = \frac{2|S \cap R|}{|S|+|R|}, \quad \text{VOE} = \frac{1-|S \cap R|}{|S \cup R|} \times 100,$$
$$\text{VD} = \frac{|S|-|R|}{|R|} \times 100.$$

Table 1 shows values of metrics with their respective standard deviation and 95% confidence interval (CI) calculated within the prescribed ROI for the SKI10 data following a 5-fold CV. μ-net showed improved VOE and, over- compared to under-estimation of volume when compared with the Imorphics [9]. Concurrent to this submission, we are also submitting our results on the testing database to the SKI10 challenge for evaluation by the organizers.

Table 2 shows these metrics along with 95% CI for 6 structures on the OAI data following a 5-fold CV. We have obtained an average Dice score of 0.849 for femoral, 0.832 for tibial and 0.785 for patellar cartilages. The Dice score for patellar cartilage was less due to absence of ground truth in few highly progressed OA cases owing to complete loss in morphology. In OAI data each of the 88 patient contribute two MR volumes (~1 year apart) which can independently fall in testing and training dataset may introduce some bias in the estimated error metrics.

Typical mid-sagittal slice of the OAI data (Fig 2(a)) with overlaid segmentation labels (Fig 2(b)) is shown in Fig 2. Over (red), under (green) and overlap (yellow) in segmented cartilage w.r.t. ground truth is shown in Fig. 2(c).

A mid-sagittal slice of the OAI data with an osteophyte (yellow arrow) at baseline and 16-months follow up is shown in Fig 3. On comparing the longitudinal data, a slight depletion of cartilage is observed in the MR slices as well as

Structures	Dice Score	VOE	VD	Avg HD
Background	0.9983 (0.9982, 0.9985)	0.3299 (0.3043, 0.3556)	0.0021 (-0.0209, 0.0250)	0.002 (0.0018, 0.0022)
Femoral Cartilage	0.849 (0.8411, 0.8570)	26.135 (24.9531, 27.3169)	2.9145 (-0.0914, 5.9205)	0.239 (0.2311, 0.2469)
Lateral Meniscus	0.8493 (0.8446, 0.8540)	26.0982 (25.4095, 26.7870)	-1.3713 (-5.8228, 3.0801)	0.2377 (0.2176, 0.2577)
Lateral Tibial Cartilage	0.8565 (0.8463, 0.8668)	24.9702 (23.4347, 26.5057)	-1.1766 (-6.5550, 4.2018)	0.2122 (0.1867, 0.2377)
Medial Meniscus	0.8018 (0.7919, 0.8118)	32.7577 (31.4692, 34.0462)	-3.7778 (-10.8962, 3.3406)	0.3864 (0.3458, 0.4269)
Medial Tibial Cartilage	0.8066 (0.7871, 0.8260)	32.1562 (29.4405, 34.8720)	-1.7158 (-6.6450, 3.2135)	0.4166 (0.3431, 0.4901)
Patellar Cartilage	0.7847 (0.7462, 0.8232)	34.1781 (29.9963, 38.3599)	3.3688 (-0.5204, 7.2579)	1.8155 (-1.4460, 5.0770)

Table 2: Metrics for on the OAI DESS studies for the six structures and background along with 95% CIs.

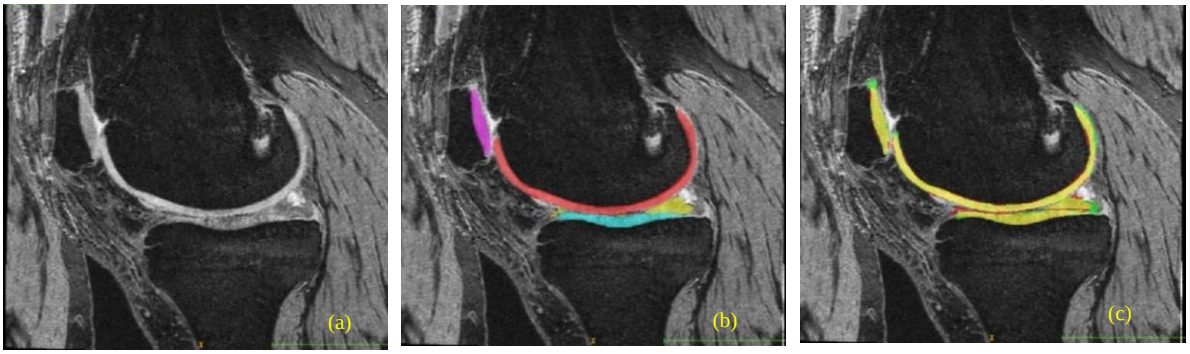

Fig 2 (a) 3D-DESS MR Slice with overlaid (b) segmentation map generated by µ-net (red: femoral cartilage, pink: patellar cartilage, green: meniscus, blue: tibial cartilage). (c) Comparison of segmentation with the ground truth (yellow: correct-segmentation, red: over-segmentation, green: under-segmentation).

in the segmentation results. Thus, highlighting the potential clinical utility of the proposed algorithm.

4. CONCLUSION

In this manuscript, we have presented µ-net, first (to the best of our knowledge) 3D-CNN for cartilage segmentation.

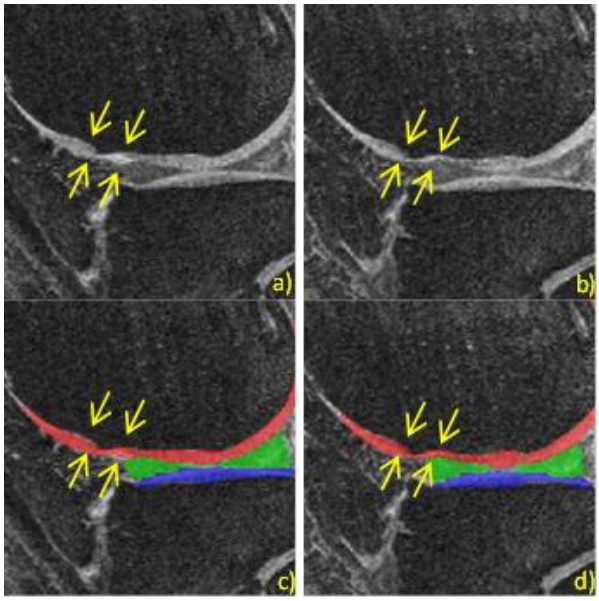

Fig. 3: a) Baseline and (b) 16 months followup mid-saggital MR Slice with segmentation on (c) baseline scan d) 16 month follow-up scan

We have shown that µ-net performed better than the current state of the art in clinically acceptable runtimes. Our Dice score measure varies from 78.5% to 85.7% for various cartilage surfaces for an input resolution of $1\times1\times1.5mm^3$. These Dice score are close to the reported mean inter-observer reproducibility of 87.7% [7]. With the increasing availability of datasets, computational resources, advances in cartilage mapping techniques such as T2, T1ρ, we believe that the proposed methods will have improved accuracy and enable automatic quantitative evaluation of knee cartilage morphology for the adoption of the quantitative MRI techniques for OA in routine clinical practice.

5. REFERENCES

[1] Y. Zhang, et al., "Epidemiology of osteoarthritis," *Clinics in geriatric medicine*, pp. 355-69, 2010.

[2] M. Cross et al., "Global burden of hip & knee osteoarthritis: estimates from the global burden of disease 2010 study," *Annals of the Rheumatic Diseases*, 2014.

[3] M.L. Gray, et al., "Toward imaging biomarkers for osteoarthritis," *Clinical orthopaedics and related research*, pp. S175-S181, 2004.

[4] C.R. Chu, et al., "Early diagnosis to enable early treatment of pre-osteoarthritis," *Arthritis research & therapy*, pp 212, 2012.

[5] X. Li, et al., "In vivo T 1ρ and T 2 mapping of articular cartilage in osteoarthritis of the knee using 3T MRI," *Osteoarthritis and cartilage*, pp. 789-797, 2007.

[6] F. Eckstein et al., "Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis," *NMR Biomed*, 2006.

[7] H. Shim, et al., "Knee Cartilage: Efficient and Reproducible Segmentation on High-Spatial-Resolution MR Images with the Semiautomated Graph-Cut Algorithm Method," *Radiology*, pp. 548-556, 2009.

[8] T. Heimann, et al., "Segmentation of knee images: A grand challenge," *MICCAI Workshop on Medical Image Analysis for the Clinic*, pp. 207-214, 2010.

[9] G. Vincent, et al., "Fully automatic segmentation of the knee joint using active appearance models," *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 224-230, 2010.

[10] A. Garcia-Garcia, et al., "A Review on Deep Learning Techniques Applied to Semantic Segmentation," *arXiv*, 2017.

[11] A. Prasoon, et al., "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," *MICCAI*, pp. 246–53, 2013.

[12] F. Milletari, et al., "V-net: Fully convolutional neural networks for volumetric medical image segmentation", *CVPR*, pp. 565-71, 2016

[13] M.C. Nevitt, et al., "The Osteoarthritis Initiative: Protocol for the Cohort Study," *National Institute of Arthritis, Musculoskeletal and Skin Diseases*, 2006. [http://oai.epi-ucsf.org]

[14] C.Y. Lee, et al. "Deeply-supervised nets," *Artificial Intelligence and Statistics*, pp. 562-570, 2015.

[15] K. He, et al., "Deep residual learning for image recognition," *CVPR*, pp. 770-78, 2016.

[16] O. Çiçek et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," *International Conference on MICCAI*, pp. 424-432, 2016.

[17] A. Raj, et al., "3D Deep Learning Based Cartilage Segmentation for Visualization and Analysis of Quantitative MRI for Osteoarthritis," *10th ISMRM Workshop on Osteoarthritis Imaging*, 2017.