Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

ALEXANDER TACK, STEFAN ZACHOW

Accurate Automated Volumetry of Cartilage of the Knee using Convolutional Neural Networks: Data from the Osteoarthritis Initiative[1]

[1]Preprint submitted to the IEEE International Symposium on Biomedical Imaging (ISBI) conference, 2019. 978-1-5386-5541-2/18/$31.00

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Accurate Automated Volumetry of Cartilage of the Knee using Convolutional Neural Networks: Data from the Osteoarthritis Initiative

Alexander Tack[1] and Stefan Zachow[1]

Zuse Institute Berlin, Berlin, Germany, tack@zib.de

Abstract. Volumetry of the cartilage of the knee, as needed for the assessment of knee osteoarthritis (KOA), is typically performed in a tedious and subjective process. We present an automated segmentation-based method for the quantification of cartilage volume by employing 3D Convolutional Neural Networks (CNNs). CNNs were trained in a supervised manner using magnetic resonance imaging data as well as cartilage volumetry readings given by clinical experts for 1378 subjects. It was shown that 3D CNNs can be employed for cartilage volumetry with an accuracy similar to expert volumetry readings. In future, accurate automated cartilage volumetry might support both, diagnosis of KOA as well as assessment of KOA progression via longitudinal analysis.

Keywords: Deep learning, imaging biomarker, radiomics, cartilage morphometry, volume assessment

1 Introduction

Knee osteoarthritis (KOA) is a degenerative disease which causes pain and decreased mobility. It affects over 40 million Europeans and the costs are estimated at 0.5% of the gross national product [1]. Articular cartilage is one of

the most relevant tissues involved in the disease process. It is recommended to employ Magnetic Resonance Imaging (MRI) for assessment of cartilage degeneration [2]. Several biomarkers derived from MRI data such as cartilage volume, cartilage thickness, and denuded cartilage area are used to assess the risk for KOA development and KOA progression [3–6]. In particular, quantification of longitudinal loss of cartilage volume shows potential for this assessment. Thus, it is desired to apply cartilage volumetry to the data of large epidemiologic studies such as the Osteoarthritis Initiative (OAI) [7], the Study of Health In Pomerania [8], or a recently performed BMBF study [9]. This, however, requires precise 3D segmentations for at least two time points. Manual cartilage segmentation is time-consuming, tedious, and results in large inter- and intra-observer variations, rendering a longitudinal analysis challenging. Moreover, Graichen et al. [10] reported absolute differences between cartilage volume analysis based on MRI and the patients' actual *in vivo* cartilage volume of 10.5% for the lateral tibial cartilage (LTC) and 11.5% for the medial tibial cartilage (MTC). The systematic difference was 3.6% and −3.1%, respectively [10].

For supervised training of automated segmentation methods, a dataset containing segmentation masks of the articular cartilage is accessible via the OAI database. This dataset was already used in previous approaches, e.g. for training k-nearest neighbour classification [11], 3D CNNs segmenting multiple small MRI subvolumes [12, 13], and 3D CNNs segmenting one larger MRI subvolume [14]. However, the given segmentation masks were generated using a semi-automated segmentation tool. For this reason, they are not sufficiently precise to train an algorithm and to evaluate its clinical value.

In this study we trained 3D CNNs employing reliable volumetry data from the OAI. Since no segmentation masks were provided together with these data, we generated our own. We utilized the OAI reference segmentations to compare

our method to other approaches (cf. [11–14]). Finally, we showed that 3D CNNs can be trained via accurate segmentation masks and volumetric readings such that the agreement between automated cartilage volumetry and clinical experts' volume readings is significantly improved compared to existing approaches.

2 Materials and Methods

Sagittal Double Echo Steady State MRI data of 1378 subjects from the OAI were utilized for our study. We employed the method of Ambellan et al. [13] for fully automated pre-segmentation of the MRI data (i.e. the tibial bone and the tibial cartilage). 3D CNNs were trained using these automatically computed masks and the effect of adding experts' cartilage volume readings as an additional criterion to the loss function was evaluated for the purpose of establishing a fully automated and accurate method to determine the volume of cartilage. Finally, the segmentation accuracy was quantified on OAI reference segmentations.

Our work yields two major contributions:
i) Training 3D CNNs using automatically computed segmentation masks. **The results outperform established methods for articular cartilage segmentation.**
ii) Adding an additional loss term considering the volume difference between automated volumetry and volumetric measures given by experts. **The results of our automated method is as accurate as the experts' volumetry.**

2.1 MRI datasets

Our methods were assessed on MRI data from the OAI:

– *Dataset Chondrometrics*: Measures of tibial cartilage volume for 1378 subjects provided by Chondrometrics (Ainring, Germany).

– *Dataset Imorphics*: Segmentations of tibial cartilage for 88 subjects (2 time points: baseline and 12-months follow-up) provided by Imorphics (Manchester, UK).

These two datasets do not share any common subjects. *Dataset Chondrometrics* consists of a larger sample size than *Dataset Imorphics* and high trust can be put into the volumetry readings performed by the clinical experts of Chondrometrics. Unfortunately, no segmentation masks were supplied by the OAI database for the subjects of this dataset.

2.2 Articular cartilage volumetry employing 3D CNNs

3D CNN architecture

3D U-Nets with an architecture similar to the 2D U-Nets as proposed by Ronneberger et al. [15] were employed (for more details see e.g. [12]). To consider memory consumption, regions of interest (ROIs) were computed such that the data was reduced to a dimension of 160×64×128 (anterior-posterior, superior-inferior, lateral-medial – cf. Fig. 1). We utilized the fact that tibial cartilage is located on the tibial plateau. Hence, a subvolume of the MRI as well as one of the corresponding segmentation mask were extracted at the superior margin of the segmented tibial bone for supervised training.

The employed 3D CNNs had five convolutional layers with two 3×3×3 convolutions per layer and an increasing number of kernels (32, 64, 128, 256, 512). The 3×3×3 convolutions were activated by ReLUs, except for the last 1×1×1 convolution in which a sigmoid function was utilized. In each layer the first convolution was followed by 10% Dropout.

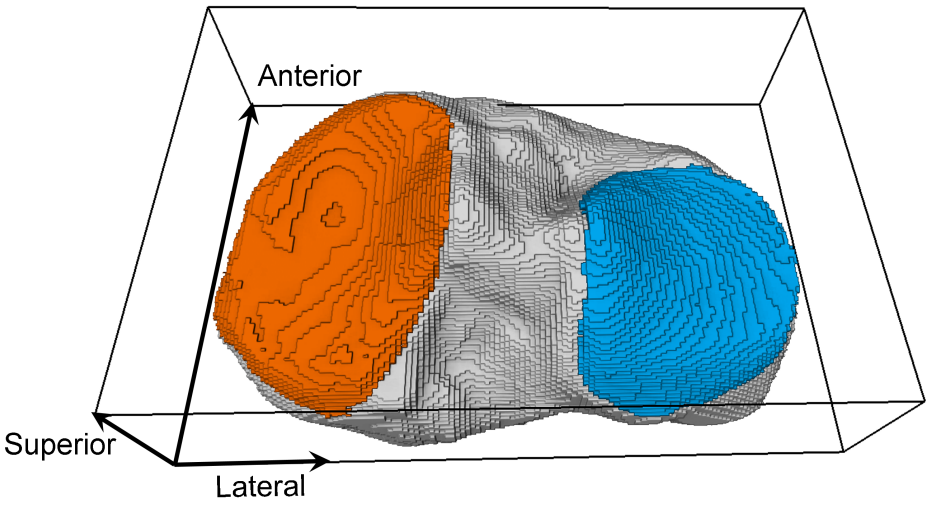
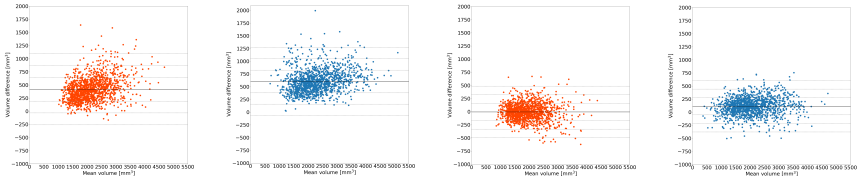


Fig. 1. Illustration of a 3D ROI (wire box). Volume rendering of segmentation masks of tibial bone (gray), medial tibial cartilage (orange), and lateral tibial cartilage (blue).

Training and Evaluation Approaches

The quality of CNNs in the context of semantic image segmentation is often measured as how well a segmentation result matches a gold standard, as for instance being given by human experts. In this study, two different approaches for such an evaluation were investigated. In a first Approach $\mathcal{A}$ we evaluated the potential of 3D CNNs trained on automatic pre-segmentations. The Dice Similarity Coefficient (DSC) was employed as a two-class loss function (class 1: MTC; class 2: LTC). We hypothesize that these CNNs are capable to learn the semantic information provided by the training data and to generalize well for the data of *Dataset Imorphics*. However, this training data was not created by clinical experts using appropriate tools. Thus, albeit an adequate agreement to the cartilage volume measures by Chondrometrics is desired, this cannot be expected employing Approach $\mathcal{A}$. For this reason, we further evaluated a second Approach $\mathcal{B}$, in which the volumetry information provided by Chondrometrics is additionally taken into account during training. Hence, in Approach $\mathcal{B}$, the



(a) Appr. $\mathcal{A}$: MTC (b) Appr. $\mathcal{A}$: LTC (c) Appr. $\mathcal{B}$: MTC (d) Appr. $\mathcal{B}$: LTC

Fig. 2. Bland-Altman plots for Approach $\mathcal{A}$ (left) and Approach $\mathcal{B}$ (right). Comparison of our automated MTC and LTC volume measures vs. those provided by Chondrometrics.

Volume Difference (VD) was considered for MTC and LTC, respectively. The VD is defined as

$$\mathrm{VD} = 100 \cdot \frac{C - \mid Y \mid}{\mid Y \mid} \ , \tag{1}$$

with C being the gold standard measures of MTC and LTC volume as provided by Chondrometrics and Y being the respective automated segmentation. The total loss function L in Approach $\mathcal{B}$ was given as the sum of DSC and VD:

$$\mathrm{L} = \mathrm{DSC} + \alpha \mathrm{VD} \ , \tag{2}$$

where VD was weighted during training by the factor $\alpha \in \mathbb{R}_+$ in order to equalize the contribution of the two summands (after each epoch α was updated). Both approaches were implemented using TensorFlow and Keras and trained using an NVIDIA Tesla P100 GPU.

3 Experiments and results

3D CNNs were trained on *Dataset Chondrometrics* using the two different loss functions as described in 2.2. For this purpose the dataset was randomly divided into two groups and both approaches were trained in a two-fold cross-validation setting employing the respective dataset only. As a means of data augmentation additional subvolumes were extracted at three random positions around the

cartilage (normal distribution: $\mu = 0$ voxel, $\sigma = 20$). This procedure resulted in 2756 training images for each group of the cross-validation split.

As shown in Table I for the data of *Dataset Imorphics*, Approach $\mathcal{A}$ yielded a segmentation accuracy of 88.02 ± 4.62 for MTC and 91.27 ± 2.33 for LTC at baseline and 87.43 ± 4.02 and 90.78 ± 2.42 at 12-months follow-up (12m). Moreover, Approach $\mathcal{A}$ resulted in a mean volume difference to the values of Chondrometrics of 22.7% ± 13.5 for MTC and 34.0% ± 22.4 for LTC (Fig. 2). Approach $\mathcal{B}$ yielded higher volume agreement on *Dataset Chondrometrics* with 0.3% ± 8.75 for MTC and 6.2% ± 9.4 for LTC (Fig. 2), but in *Dataset Imorphics* the segmentation accuracy decreased slightly to 82.85 ± 5.53 for MTC and 86.11 ± 4.37 for LTC at baseline and 82.27 ± 5.80 and 85.83 ± 4.30 at 12m.

In comparison, the method proposed by Dam et al. over-estimated the cartilage volume on average by 4% for MTC and 14% for LTC compared to the measures by Chondrometrics [11]. We confirmed the over-segmentation via the method of Ambellan et al., where the Imorphics data acted as gold standard for training and no volumetric loss was used. Applying the method of Ambellan et al. to the data of *Dataset Chondrometrics*, the volume error was 19.0% and 27.6% for MTC and LTC, respectively. Compared to Dam et al. and Ambellan et al., Approach $\mathcal{B}$ achieved a similar segmentation accuracy w.r.t. the DSC on *Dataset Imorphics* – but yielded volumes clearly closer to the ones provided by the clinical experts of Chondrometrics.

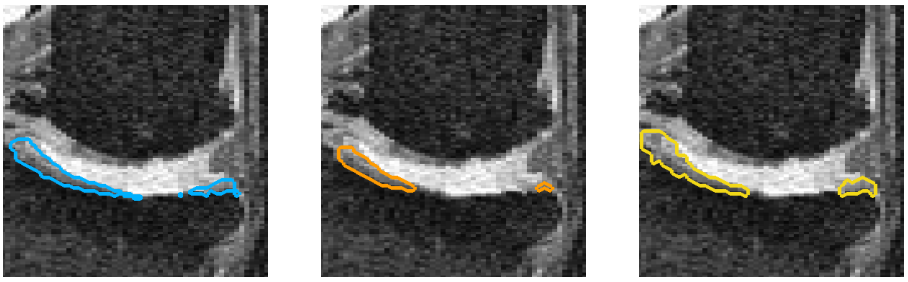
Complete volumetric analysis of the tibial cartilage took on average approx. 5s per knee on a workstation with NVIDIA GTX 1080 Ti GPU.

4 Discussion and conclusion

Employing Approach $\mathcal{A}$, the segmentation accuracy outperformed previous methods as measured with the DSC for data of *Dataset Imorphics*. We hypoth-

Table I. Results of different methods for the automated segmentation of MTC and LTC in *Dataset Imorphics* (baseline and 12-months follow-up time point).

	baseline	
	MTC	LTC
Dam et al. [11]	81.20 ± 5.50	86.60 ± 3.40
Raj et al. [14]	80.66	85.65
Tack et al. [12]	85.13 ± 10.5	90.23 ± 4.64
Ambellan et al. [13]	86.10 ± 5.33	90.40 ± 2.42
Approach $\mathcal{B}$	82.85 ± 5.53	86.11 ± 4.37
Approach $\mathcal{A}$	**88.02 ± 4.62**	**91.27 ± 2.33**
	12m	
	MTC	LTC
Dam et al. [11]	—	—
Raj et al. [14]	—	—
Tack et al. [12]	85.86 ± 5.03	90.20 ± 2.64
Ambellan et al. [13]	85.80 ± 5.00	89.10 ± 2.41
Approach $\mathcal{B}$	82.27 ± 5.80	85.83 ± 4.30
Approach $\mathcal{A}$	**87.43 ± 4.02**	**90.78 ± 2.42**



(a) Approach $\mathcal{A}$ (b) Approach $\mathcal{B}$ (c) Imorphics

Fig. 3. Coronal view of the medial compartment of OAI patient Id #9892765: Results of Approach $\mathcal{A}$ (a), Approach $\mathcal{B}$ (b), and reference segmentation of Imorphics (c).

esize that the increase in accuracy was due to both, the full 3D character of our CNN as well as the additional amount of training data. Although the segmentation accuracy on *Dataset Imorphics* illustrates the validity of 3D CNNs for the task of segmentation, over-estimation of the cartilage volume compared to Chondrometrics is severe, hence, limiting an application for automated cartilage analysis.

Approach $\mathcal{B}$ yielded a significantly higher agreement in volumetry on *Dataset Chondrometrics*, with a slight decrease of the segmentation accuracy for the data of *Dataset Imorphics*. As shown in Fig. 3 the specificity is slightly increased in areas of degenerated cartilage, which is hard to differentiate from healthy tissue.

In future, we will apply our accurate volumetry method for the entire OAI database, to provide segmentation masks and volumetry data, that can be used as a gold standard for longitudinal assessment of OA progression as well as for the development of novel methods.

References

1. P.G. Conaghan, M. Kloppenburg, G. Schett, J.W.J. Bijlsma, et al. Osteoarthritis research priorities: a report from a EULAR ad hoc expert committee. *Annals of the Rheumatic Diseases*, pages 1–4, 2014.
2. P.G. Conaghan, D.J. Hunter, J.F. Maillefert, W.M. Reichmann, and E. Losina. Summary and recommendations of the OARSI FDA osteoarthritis assessment of structural change working group. *Osteoarthritis and Cartilage*, 19(5):606–610, 2011.
3. F. Eckstein, D. Burstein, and T.M. Link. Quantitative MRI of cartilage and bone: Degenerative changes in osteoarthritis. *NMR in Biomedicine*, 19(7):822–854, 2006.
4. F. Eckstein, F. Cicuttini, J.-P. Raynauld, J.C. Waterton, and C. Peterfy. Magnetic Resonance Imaging (MRI) of articular cartilage in knee osteoarthritis (OA): Morphological assessment. *Osteoarthritis and Cartilage*, 14:46–75, 2006.
5. F. Eckstein, S. Maschek, W. Wirth, M. Hudelmaier, et al. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative Progression Subcohort-assosiation with sex, Body Mass Index, symptoms, and radiographic OA status. *Annals of the Rheumatic Diseases*, 68(5):1–16, 2008.

6. J. Hohe, G. Ateshian, M. Reiser, K.-H. Englmeier, and F. Eckstein. Surface size, curvature analysis, and assessment of knee joint incongruity with MRI in vivo. *Magnetic Resonance in Medicine*, 47(3):554–561, 2002.

7. OAI. https://oai.nih.gov, 2019.

8. U. John, E. Hensel, J. Lüdemann, M. Piek, et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Sozial-und Präventivmedizin*, 46(3):186–194, 2001.

9. OVERLOAD-PrevOP: Verständnis und Prävention der Progression der primären Osteoarthrose. https://gesundheitsforschung-bmbf.de/de/OVERLOAD-PrevOP-Praevention-Progression-primaere-Osteoarthrose.php, 2019.

10. H. Graichen, R. v. Eisenhart-Rothe, T. Vogl, K.-H. Englmeier, and F. Eckstein. Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging: Technical validation for use in analysis of cartilage volume and further morphologic parameters. *Arthritis & Rheumatism*, 50(3):811–816, 2004.

11. E.B. Dam, M. Lillholm, J. Marques, and M. Nielsen. Automatic segmentation of high-and low-field knee MRIs using knee image quantification with data from the Osteoarthritis Initiative. *Journal of Medical imaging*, 2(2):024001, 2015.

12. A. Tack, A. Mukhopadhyay, and S. Zachow. Knee menisci segmentation using convolutional neural networks: Data from the Osteoarthritis Initiative. *Osteoarthritis and Cartilage*, 26(5):680–688, 2018.

13. F. Ambellan, A. Tack, M. Ehlke, and S. Zachow. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. In *International Conference on Medical Imaging with Deep Learning*, 2018.

14. A. Raj, S. Vishwanathan, B. Ajani, K. Krishnan, and H. Agarwal. Automatic knee cartilage segmentation using fully volumetric convolutional neural networks for evaluation of osteoarthritis. In *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 851–854, 2018.

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