

On Nonparametric Markov Random Field Estimation for Fast Automatic Segmentation of MRI Knee Data^{*}

Filip Korč, David Schneider and Wolfgang Förstner

Dept. of Photogrammetry, University of Bonn,
Nussallee 15, 53115 Bonn, Germany
{[filip.korc](mailto:filip.korc@uni-bonn.de), [schneide](mailto:schneide@uni-bonn.de), [wf](mailto:wf@uni-bonn.de)}@uni-bonn.de
<http://www.ipb.uni-bonn.de/>

Abstract. We present a fast automatic reproducible method for 3d semantic segmentation of magnetic resonance images of the knee. We formulate a single global model that allows to jointly segment all classes. The model estimation was performed automatically without manual interaction and parameter tuning. The segmentation of a magnetic resonance image with 11 Mio voxels took approximately one minute. Our labeling results by far do not reach the performance of complex state of the art approaches designed to produce clinically relevant results. Our results could potentially be useful for rough visualization or initialization of computationally demanding methods. Our main contribution is to provide insights in possible strategies when employing global statistical models.

Keywords: Bone, cartilage, knee, magnetic resonance imaging (MRI), 3d semantic segmentation, Markov random field (MRF), nonparametric density estimation.

1 Introduction

We present a fast automatic reproducible method for 3d semantic segmentation of magnetic resonance (MR) images of the knee. We formulate a single global model that allows to jointly segment all classes. The model adaptation was performed automatically without manual interaction and parameter tuning. The segmentation of an MR image with 11 Mio voxels took approximately one minute.

From a broad perspective in this work we investigate the potential of Markov random field (MRF) models [6] in this context. More specifically, we investigate to what extent can we in this context eliminate the need for solving an involved task of learning unknown model parameters.

^{*} In Proceedings of the 4th *Medical Image Analysis for the Clinic - A Grand Challenge* workshop, MICCAI 2010.

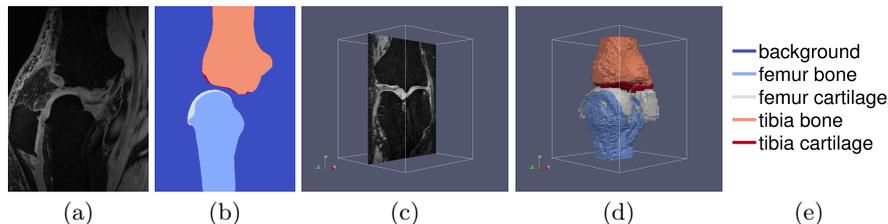


Fig. 1. (a) 2d slice of a training MR image. (b) Manual segmentation of the MR image in (a), where we show the corresponding 2d slice. (c) 2d slice of a test MR image shown in the outlined 3d MR image volume. (d) Our automatic segmentation of the test MR image in (c), where we show the corresponding 3d view with transparent background class labels. (e) Class label color code used in (b) and (d).

Markov random fields are global models that in general pose a problem of estimating a complex probability distribution. A widely adopted approach is to first formulate a model of specific functional form governed by some number of unknown parameters and subsequently to learn these parameters from a training data set. Learning unknown parameters of a global probabilistic model is a task that can in principle be solved by a Monte Carlo simulation [8] or, alternatively, by methods of variational inference [7]. However, the task involves solving an NP-hard problem and, hence, in practice both approaches in general inevitably lead to approximations. Even though much work has recently been done in both areas, approximate parameter learning remains a difficult task.

We pose the following practical question. To what extent can parameter learning be avoided? In this work we completely eliminate the need for solving a difficult learning problem by designing a nonparametric model and reducing the model adaptation to a simple histogram based nonparametric density estimation. Let us note that while large volumes of magnetic resonance imaging (MRI) data would further challenge the efficiency of approximate parameter learning approaches, for nonparametric density estimation techniques, large amount of data is of advantage.

In this work we focus on a generative model. Such model, as opposed to a discriminative model, poses in general a more complex problem, however, it has the advantage of allowing not only to find an unknown labeling, but a generative model also provides means to compute labeling of incomplete data, it can be used to sample an MR image using a Monte Carlo simulation and quite importantly can be combined with additional priors or be integrated in a superior reasoning system. Disadvantage of the model is its lower discriminative power, relative to a pure discriminative model. In case we would be mainly interested in finding an unknown labeling, it would be possible to adopt a similar approach and sacrifice the estimation of the generative power in favor of a better estimate of a pure discriminative model.

2 Related Work

In clinical studies, a manual or semi-automated 2d slice by slice segmentation is a prevalent approach, leading to possibly several hours of trained radiologist’s manual slice annotation. In the context of medical image analysis, number of both semi-automatic and automatic approaches have been proposed. These approaches range from 2d slice by slice methods, or methods combining this approach with 3d image registration, to the methods segmenting directly in 3d. In our review we focus on current approaches that are automatic and that segment directly in 3d.

Current state of the art approaches [2][3][1] are complex problem specific methods comprised of multiple locally acting computational steps performed sequentially. We formulate a simple reproducible global model that in one step allows to jointly segment all classes. Our work is in this regard related to the work in [4]. An atlas based prior information has previously been used both in the form of one particular manual segmentation [3] or in the form of a statistical atlas [4] that was derived from multiple manual segmentations. Our model utilizes a spatially variant label prior that we derive from multiple manual segmentations and that can be viewed as a form of a statistical atlas. Segmentation time in [1] is in the order of tens of minutes. Cartilage segmentation in [3] takes approximately 15 minutes. We segment an MR image with 11 Mio voxels in approximately one minute, which is comparable to [4]. Our approach is related to the work in [4] both conceptually and computationally. Broadly they also adopt MRF model and also employ the optimization algorithm proposed in [5]. They are interested in finding patella cartilage and pose the problem as registering a deformable statistical atlas to the observed data. We are interested in a multi-label problem that we pose as a semantic segmentation task assuming that prior label information can be captured in a spatially variant label prior and spatially invariant label smoothness prior. We currently study the relation of our work to the approach in [4]. Regarding the accuracy, our labeling results currently by far do not reach the state of the art performance of complex problem-specific approaches in [2][3][1][4].

3 Task and Model Requirements

We are faced with a multi-label semantic segmentation problem, where we wish to identify each voxel of an observed and previously unseen MR image with the femur, the femoral cartilage, the tibia, the tibial cartilage or with the background class label.

We assume that complex MRI measurement yields noisy observation and, hence, to account for the expected interscan variability we require our data model to be statistical. It is our goal, among others, to distinguish between the femur, the tibia and the patella. Let us maintain that we wish to identify the latter with the background class. Inspection of Fig. 1(a) and Fig. 1(b) reveals that these bones appear as large regions with pronounced boundaries. Hence,

we assume that data provides us with a hint in regard to the spatial extent of these regions. Nevertheless further inspection reveals that all bones appear almost identical and also very similar to the area outside of the knee in the top left corner in Fig. 1(a). Hence, we assume that MR data only provides us with a hint of a class subset. On the other hand we realize that MRI is a controlled measurement. Hence, we assume that the location in the MR image provides us with a hint in regard to the class label. Nevertheless we expect variability in the knee placement and in the knee itself and assume thus that at the same time location in the MR image does not provide a hint of region boundaries. This motivates the model requirement to provide means to statistically capture spatially variant label prior information and to combine this information with the statistical data model. Background and bones are large homogeneous regions, see Fig. 1(b), and, hence, we require our model to prefer smooth segmentation results.

The femoral and the tibial cartilage appear to challenge all model requirements. Cartilage lacks both distinct appearance and distinct boundaries with neighboring soft tissues and represents thus challenge to the data model. Being realized by a thin layer of tissue, cartilage represents a challenge for both the spatially variant label prior, that should be well localized, and the smoothness prior, that should not oversmooth the resulting segmentation. Hence, we require the model to well balance all requirements.

4 Nonparametric Markov Random Field Model

In this section we derive a global statistical model that relates the unobservable class labels to an observed MR image. In other words, let a vector \mathbf{y} denote an MR image, where $\mathbf{y} = (y_i)_{i \in S}$, the scalar y_i is the intensity of the i th voxel, and S is the set of the MR voxel indices. Let the class labels assigned to the MR voxels be given by a vector $\mathbf{x} = (x_i)_{i \in S}$. We have $x_i \in \{0, 1, 2, 3, 4\}$, where the label set encodes the set {background, femur, femoral cartilage, tibia, tibial cartilage} of classes of our multi-label semantic segmentation problem. We refer to the vector \mathbf{x} as label configuration and sometimes abbreviate label configuration as labeling. In the considered generative framework we relate an unobservable class label configuration \mathbf{x} to an observed MR image \mathbf{y} in terms of the joint probability $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} | \mathbf{x})p(\mathbf{x})$.

We in common way simplify complex data generation process $p(\mathbf{y} | \mathbf{x})$ as $p(\mathbf{y} | \mathbf{x}) \equiv \prod_i p(y_i | x_i)$. This step is needed to arrive at a computationally tractable generative model. On one hand, this means to give up discriminative power of features derived from larger image regions and to resort to using sole intensity values as features. On the other hand, the resulting simplicity allows to estimate the model component accurately from training data. We note that the functional form of the term itself is spatially invariant. Nevertheless let us maintain that the value of this term is spatially variant through the dependence on the data. We thus refer to this term as spatially invariant data term.

We model the class labeling prior $p(\mathbf{x}) \equiv p_{MRF}(\mathbf{x}) \prod_i p_i(x_i)$ as a product of a spatially variant label prior $p_i(x_i)$ and spatially invariant smoothness prior $p_{MRF}(\mathbf{x})$. We model the latter term $p_{MRF}(\mathbf{x}) \propto \exp\{\prod_{\{i,i'\}} \delta(x_i - x_{i'})\}$ as a standard Potts model [6][8] Markov random field defined on a 3-dimensional (3d) grid with 6-neighborhood. Here $\{i, i'\}$ denotes pair of neighboring voxels.

The modeled joint probability can now be expressed as

$$p(\mathbf{x}, \mathbf{y}) = p_{MRF}(\mathbf{x}) \prod_i p(y_i | x_i) p_i(x_i) \quad (1)$$

5 Model Adaptation by Nonparametric Estimation

We adopt a histogram based nonparametric density estimation approach and model the spatially invariant class-conditional intensity probability $p(y_i | x_i)$,

$$p(y_i | x_i) \equiv \hat{p}(y_i | x_i) \quad (2)$$

as an empirical probability estimate $\hat{p}(y_i | x_i)$, i.e., as a class-specific frequency of an intensity value in training data divided by the total number of its observations.

Similarly, the spatially variant class label prior distribution $p_i(x_i)$,

$$p_i(x_i) \equiv \hat{p}_i(x_i) \quad (3)$$

is an empirical probability estimate $\hat{p}_i(x_i)$ of a class label at a specific location.

Unlike nearest-neighbor based nonparametric methods and kernel based nonparametric methods, the adopted histogram based method has the advantage of allowing to discard the training data once the histograms have been computed. This is especially convenient in a situation involving large MRI data volumes. In addition, this approach further allows for a convenient model update in case the training data arrives sequentially.

6 Inference of Unknown Labeling

For the statistical model formulated in Eq. (1), for its estimate in Eq. (2), for its estimate in Eq. (3) and for a given and previously unseen MR image \mathbf{y} we are interested in finding a label configuration \mathbf{x} that maximizes the probability in Eq. (1). To do so, we express the model in Eq. (1) in an equivalent form of a Boltzmann distribution $p(\mathbf{x}, \mathbf{y}) \propto \exp\{-E(\mathbf{x}, \mathbf{y})\}$ and pose the problem as finding a minimum of the so called energy function

$$E(\mathbf{x}, \mathbf{y}) = - \sum_i \log \hat{p}_i(x_i) - \sum_i \log \hat{p}(y_i | x_i) + \sum_{\{i,i'\}} \delta(x_i - x_{i'})$$

In general, this problem is very difficult to solve exactly. Recently, algorithms based on graph cuts, loopy belief propagation and convex relaxation have proven to be efficient in finding good approximate solutions to the above problem. In our experiments we adopted the latter proposed in [5].

7 Experiments, Results and Discussion

In our experiments, we use training data to train our model and we use test data to evaluate the model. Training data comprised 60 MR images, see Fig. 1(a), and of 60 corresponding manual segmentations, see Fig. 1(b). In our work we refer to the result of manual segmentation as ground truth label configuration. Test data consisted of 40 MR images, see Fig. 1(c). Test ground truth labelings were unknown to us during experiments. Evaluation of our segmentation results, see Fig. 1(d), that involved the unknown test ground truth labelings, was performed by a third party. In total, there were 100 MR images available to us. MR image size ranged from 8 to 16 Mio voxels, where typical size was 11 Mio voxels. Typical MR image resolution was roughly $300 \times 350 \times 100$ voxels. MR image spatial resolution was $0.4 \times 0.4 \times 1$ mm. Hence, typical MR image volume size was roughly $120 \times 140 \times 100$ mm. In our experiments, we normalized the intensity values, in each MR image individually to $[0, 1]$ by finding the maximum intensity and by dividing each intensity by this value. We did not correct the bias by subtracting the minimum intensity from each intensity value.

We resort to using intensity values as features. To what extent are intensity values discriminative for our problem? What is the interscan variability of the intensities? To answer these questions we used training data to compute class-specific histograms of intensity values, see Fig. 2(a). In Fig. 2(b) we show means of single MR image frequencies together with the percentile ranges. We observe that background class intensities cover almost the whole intensity range. Further, we have observed that the femoral and the tibial cartilage class intensities cover wide intensity range and that the distributions are very similar. The femur and the tibia intensity distributions are more localized, however, appear almost identical. In Fig. 2(b) we observe that per MR image distributions can deviate considerably from each other. We conclude that the adopted features are simple on one hand, however, on the other hand appear to be weak in terms of discriminative power. For this reason, the spatially invariant data term, that is based on the class conditional distribution of intensity values shown in Fig. 2(c), is a weak component of our model. This is further confirmed in Fig. 4(b).

To what extent is the spatially variant label prior derived from unregistered manual segmentations informative? Slices of the spatially variant label prior volume with the resolution of $60 \times 60 \times 60$ are shown in Fig. 3. We show class-specific slices and also a max prior probability slice, where white means preference for a particular label and where gray area means weak preference for a particular class label. What is the role of the spatially variant label prior? This is illustrated in Fig. 4(c)(e)(f). Here we observe how prior alignment is corrected by the data term. Fig. 4(d)(f) further illustrates the role and the impact of the spatially invariant label smoothness prior.

Further qualitative and quantitative results using the full model, where a coarse spatially variant label prior with the resolution of $30 \times 30 \times 10$ was used, can be found in respectively Fig. 5 and Tab. 1. Here we observe strong block artifacts even though the segmentation was computed on the full resolution. Comparison to the result in Fig. 4(f), where a finer $60 \times 60 \times 60$ spatially variant

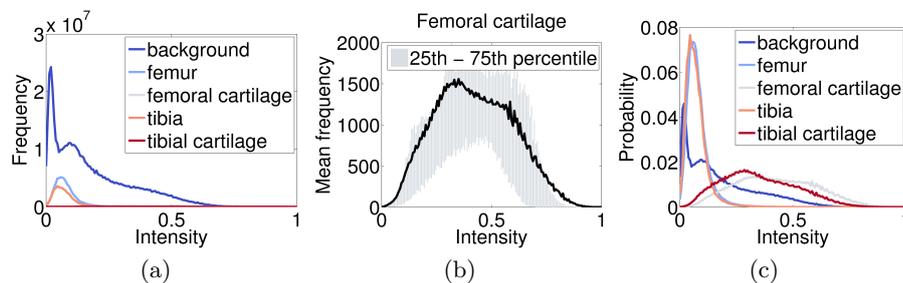


Fig. 2. Single point on a function graph shows: (a) Class-specific frequency of an intensity in training data, (b) mean with percentile range of frequencies computed per training MR image, (c) class-conditional probability of an intensity computed from the frequency in (a).

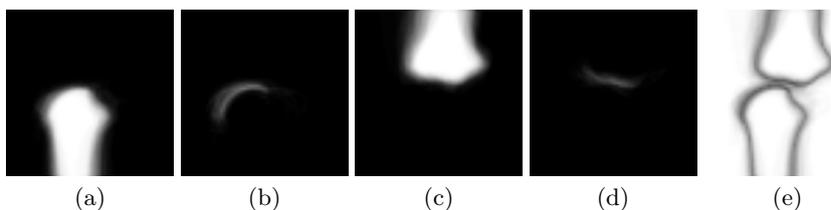


Fig. 3. Slices of spatially variant label prior. Probability values of (a) femur, (b) femoral cartilage, (c) tibia, (d) tibial cartilage. (e) Maximum prior probability values. We show probability of 1 in white, probability of 0 in black and other probability values in gray.

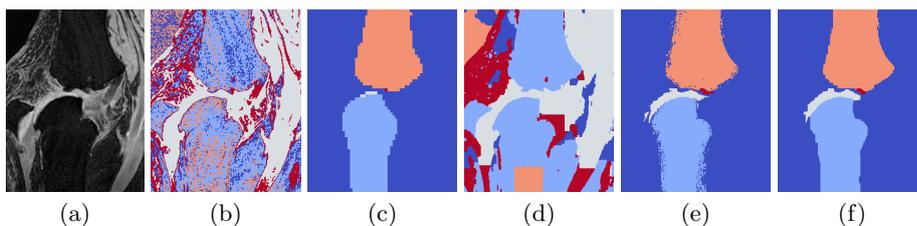


Fig. 4. (a) 2d slice of the test MR image in Fig. 1(c). Model components: 2d slice of labeling computed only with (b) spatially invariant data term, (c) spatially variant label prior, (d) combination of spatially invariant data term and spatially invariant label smoothness prior, (e) combination of spatially invariant data term and spatially variant label prior. (f) 2d slice of the full model based labeling in Fig. 1(d).

label prior was used, reveals that this is mainly a consequence of the coarse spatially variant label prior used here.

8 Conclusion

We discussed the potential of a fast automatic reproducible method for 3d semantic segmentation of magnetic resonance images of the knee. Our labeling results by far do not reach the state of the art performance of complex approaches specifically designed to produce clinically relevant results. We illustrated to what extent can an involved learning problem be avoided, while still obtaining results potentially useful for rough visualization, see Fig. 1(d), or for initialization of computationally demanding methods. Our main contribution is to provide insights in possible strategies when employing global statistical models.

Acknowledgment

We wish to thank our anonymous reviewers for pointing us to the literature and for making valuable comments in regard to the topic presentation.

References

1. P. Dodin, J. Pelletier, J. Martel-Pelletier, and F. Abram. Automatic human knee cartilage segmentation from 3d magnetic resonance images. *Biomedical Engineering*, 2010. Accepted for publication.
2. J. Folkesson, E.B. Dam, O.F. Olsen, P.C. Pettersen, and C. Christiansen. Segmenting articular cartilage automatically using a voxel classification approach. *Medical Imaging*, 26:106 – 115, 2007.
3. J. Fripp, S. Crozier, S.K. Warfield, and S. Ourselin. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. *Medical Imaging*, 29:55–64, 2010.
4. B. Glocker, N. Komodakis, N. Paragios, D. Glasser, G. Tziritas, and N. Navab. Primal/dual linear programming and statistical atlases for cartilage segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2007.
5. N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *International Conference on Computer Vision*, 2007.
6. S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 3rd edition, 2009.
7. M. J. Wainwright and M. I. Jordan. *Graphical models, exponential families, and variational inference*. Foundations and Trends in Machine Learning. 2008.
8. G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer, 2nd edition, 2006.

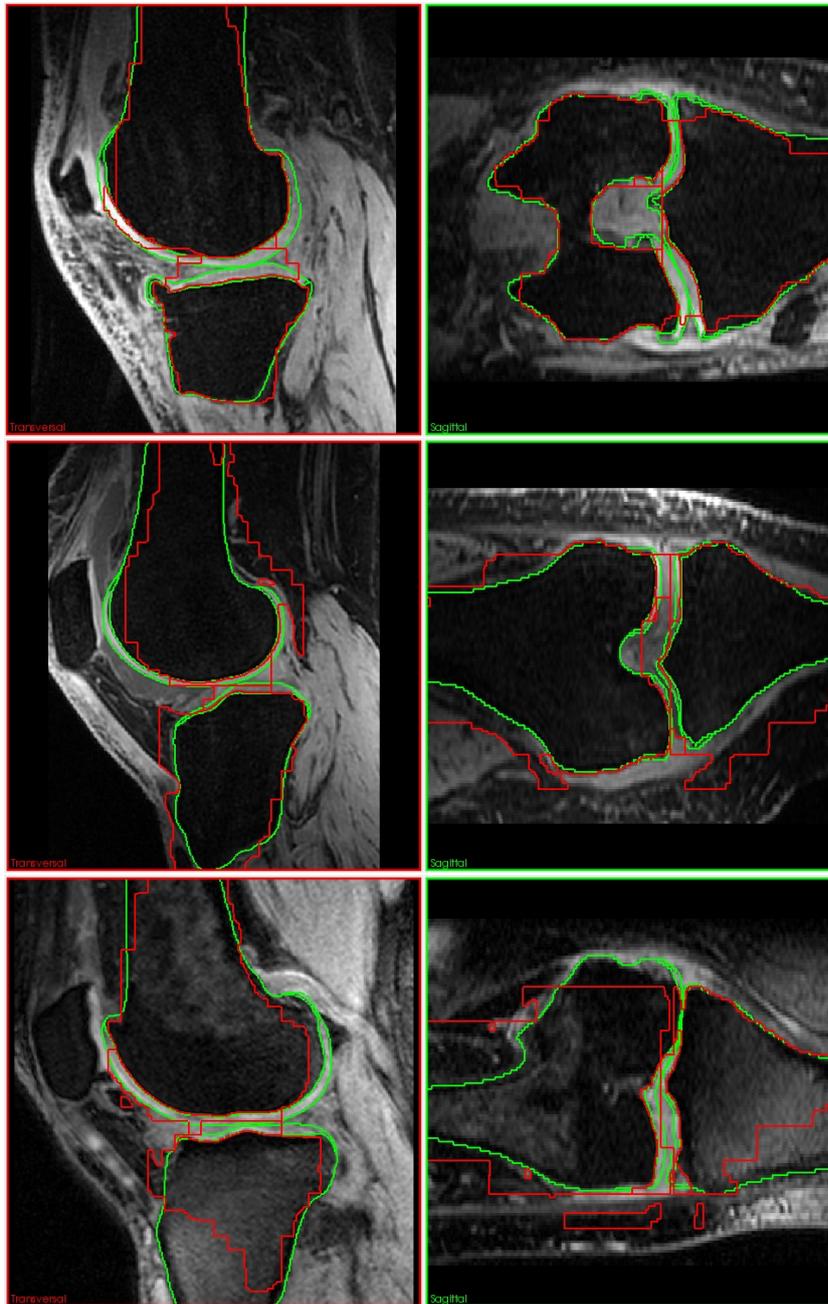


Fig. 5. Different views on selected test cases 8 (top), 16 (center), and 31 (bottom). The outline of the reference segmentation is displayed in green, the outline of the automatic method described in this paper in red.

| Img | Femur bone | | | Tibia bone | | | Fem. cartilage | | | Tibial cartilage | | | Total Score |
|-----|------------|------------|---------|------------|------------|---------|----------------|------------|----------|------------------|------------|----------|-------------|
| | AvgD | RMSD | Scr | AvgD | RMSD | Scr | VOE | VD | Scr | VOE | VD | Scr | |
| | [mm] | [mm] | | [mm] | [mm] | | [%] | [%] | | [%] | [%] | | |
| 1 | 2.27 | 3.14 | 0 | 2.31 | 3.52 | 0 | 50.8 | 89.2 | 31 | 50.3 | -25.8 | 36 | 16.9 |
| 2 | 2.40 | 3.20 | 0 | 3.55 | 4.55 | 0 | 90.0 | -89.6 | 17 | 62.4 | 10.2 | 59 | 19.1 |
| 3 | 2.91 | 4.29 | 0 | 3.95 | 5.40 | 0 | 56.6 | -8.4 | 64 | 68.8 | 93.3 | 25 | 22.3 |
| 4 | 2.77 | 4.24 | 0 | 3.05 | 5.13 | 0 | 80.3 | -74.2 | 21 | 59.3 | -28.3 | 28 | 12.3 |
| 5 | 2.60 | 3.37 | 0 | 3.73 | 6.56 | 0 | 75.4 | -67.6 | 22 | 66.2 | 14.3 | 51 | 18.3 |
| 6 | 2.04 | 2.90 | 3 | 1.51 | 2.16 | 6 | 55.2 | 38.5 | 30 | 52.3 | 18.1 | 49 | 22.1 |
| 7 | 3.16 | 4.61 | 0 | 3.41 | 4.91 | 0 | 76.2 | -65.7 | 22 | 65.2 | 79.9 | 26 | 12.1 |
| 8 | 2.22 | 3.24 | 0 | 2.51 | 3.81 | 0 | 60.2 | -36.1 | 28 | 57.2 | 95.3 | 29 | 14.3 |
| 9 | 3.31 | 4.49 | 0 | 3.32 | 4.81 | 0 | 60.9 | -12.0 | 57 | 61.3 | 68.8 | 28 | 21.0 |
| 10 | 2.68 | 3.63 | 0 | 2.97 | 3.95 | 0 | 43.5 | -25.2 | 40 | 49.4 | 82.2 | 32 | 17.9 |
| 11 | 2.99 | 4.20 | 0 | 2.86 | 4.07 | 0 | 55.8 | 62.2 | 30 | 56.6 | 105.5 | 29 | 14.7 |
| 12 | 3.19 | 4.93 | 0 | 3.44 | 5.23 | 0 | 70.7 | -42.7 | 24 | 68.0 | 50.1 | 25 | 12.3 |
| 13 | 2.70 | 3.66 | 0 | 2.21 | 3.27 | 0 | 66.1 | 136.2 | 26 | 89.4 | -24.9 | 24 | 12.3 |
| 14 | 2.57 | 3.40 | 0 | 2.21 | 3.21 | 0 | 61.4 | 131.5 | 28 | 68.4 | -39.5 | 25 | 13.1 |
| 15 | 2.88 | 4.20 | 0 | 2.76 | 3.80 | 0 | 72.6 | -66.4 | 23 | 64.2 | 105.4 | 27 | 12.5 |
| 16 | 3.73 | 5.39 | 0 | 3.86 | 6.05 | 0 | 89.1 | -70.7 | 17 | 71.7 | -39.8 | 24 | 10.3 |
| 17 | 2.92 | 3.85 | 0 | 3.09 | 4.78 | 0 | 86.3 | -77.9 | 18 | 76.1 | -37.7 | 22 | 10.2 |
| 18 | 2.09 | 2.97 | 2 | 2.58 | 4.43 | 0 | 55.7 | 106.9 | 30 | 79.8 | -73.2 | 21 | 13.0 |
| 19 | 1.93 | 2.84 | 4 | 1.55 | 2.37 | 2 | 51.9 | 72.9 | 31 | 49.3 | 20.0 | 47 | 21.0 |
| 20 | 3.24 | 4.60 | 0 | 2.65 | 3.68 | 0 | 74.8 | -52.3 | 23 | 51.1 | 70.5 | 31 | 13.5 |
| 21 | 2.14 | 3.21 | 0 | 3.30 | 5.60 | 0 | 74.1 | -31.6 | 23 | 65.0 | -25.6 | 31 | 13.5 |
| 22 | 2.40 | 3.57 | 0 | 3.39 | 5.16 | 0 | 59.5 | 93.1 | 28 | 82.6 | -67.4 | 20 | 12.0 |
| 23 | 2.76 | 3.96 | 0 | 4.14 | 5.90 | 0 | 53.7 | -6.6 | 69 | 64.1 | 64.2 | 27 | 23.8 |
| 24 | 2.38 | 3.61 | 0 | 2.76 | 3.93 | 0 | 45.7 | 52.0 | 33 | 61.8 | -22.1 | 39 | 18.0 |
| 25 | 3.07 | 4.68 | 0 | 3.53 | 5.07 | 0 | 44.2 | 62.9 | 34 | 44.8 | 41.7 | 34 | 16.9 |
| 26 | 3.36 | 4.34 | 0 | 4.11 | 7.02 | 0 | 93.2 | -93.1 | 16 | 82.5 | -46.1 | 20 | 8.9 |
| 27 | 2.46 | 3.40 | 0 | 2.78 | 3.77 | 0 | 77.6 | -61.0 | 22 | 50.5 | 44.6 | 32 | 13.3 |
| 28 | 4.86 | 7.21 | 0 | 2.76 | 3.89 | 0 | 68.4 | 39.4 | 25 | 50.5 | 31.7 | 32 | 14.1 |
| 29 | 2.09 | 2.73 | 6 | 5.06 | 6.77 | 0 | 76.1 | -62.1 | 22 | 89.6 | -58.9 | 17 | 11.3 |
| 30 | 2.64 | 3.93 | 0 | 2.07 | 3.70 | 0 | 53.0 | 33.0 | 31 | 64.7 | 24.3 | 34 | 16.1 |
| 31 | 4.74 | 6.90 | 0 | 2.24 | 3.61 | 0 | 63.6 | -36.8 | 27 | 61.6 | 40.6 | 28 | 13.6 |
| 32 | 2.04 | 2.89 | 3 | 1.54 | 2.30 | 4 | 53.3 | 51.2 | 31 | 61.7 | 18.3 | 45 | 20.6 |
| 33 | 2.64 | 4.17 | 0 | 1.67 | 2.62 | 0 | 60.8 | 27.9 | 29 | 59.5 | 111.5 | 28 | 14.2 |
| 34 | 2.51 | 3.53 | 0 | 2.14 | 3.20 | 0 | 84.3 | -79.3 | 19 | 64.4 | 93.7 | 26 | 11.4 |
| 35 | 2.51 | 3.78 | 0 | 2.16 | 3.24 | 0 | 61.5 | -40.4 | 28 | 52.0 | -1.2 | 79 | 26.6 |
| 36 | 1.94 | 3.22 | 0 | 1.81 | 2.72 | 0 | 47.6 | 80.8 | 33 | 72.5 | -57.6 | 24 | 14.0 |
| 37 | 5.14 | 6.92 | 0 | 5.94 | 8.22 | 0 | 80.8 | 8.7 | 55 | 100.0 | -100.0 | 13 | 17.1 |
| 38 | 2.94 | 4.34 | 0 | 2.14 | 3.26 | 0 | 53.8 | 4.7 | 72 | 57.9 | 129.1 | 29 | 25.2 |
| 39 | 4.35 | 5.53 | 0 | 5.47 | 7.23 | 0 | 95.9 | -93.6 | 15 | 84.2 | -41.3 | 19 | 8.5 |
| 40 | 4.39 | 6.82 | 0 | 5.32 | 8.77 | 0 | 52.5 | 28.7 | 31 | 65.6 | 75.8 | 26 | 14.2 |
| Avg | 2.90 | 4.15 | 0 | 3.05 | 4.54 | 0 | 65.8 | -1.8 | 31 | 65.1 | 20.0 | 31 | 15.6 |
| | ± 0.81 | ± 1.17 | ± 1 | ± 1.09 | ± 1.59 | ± 1 | ± 14.4 | ± 66.9 | ± 14 | ± 12.7 | ± 60.7 | ± 12 | ± 4.5 |

Table 1. Results of the comparison metrics and scores for all 40 test cases. AvgD and RMSD are the average and RMS surface distance, respectively, VOE is the volumetric overlap error and VD indicates the volumetric difference.