MRI Breast Segmentation using Unsupervised Neural Networks for Biomechanical Models

S. Said[a], M. Meyling[a], R. Huguenot[a], M. Horning[a], P. Clauser[b], N.V. Ruiter[a], P.A.T. Baltzer[b], and T. Hopp[a]

[a]Karlsruhe Institute of Technology (KIT), Institute for Data Processing and Electronics, Karlsruhe, Germany
[b]Medical University of Vienna, Department of Biomedical Imaging and Image-guided Therapy, Vienna, Austria

ABSTRACT

In multimodal diagnosis for early breast cancer detection, spatial alignment by means of image registration is an important task. We develop patient-specific biomechanical models of the breast, for which one of the challenges is automatic segmentation for magnetic resonance imaging (MRI) of the breast. In this paper, we propose a novel method using unsupervised neural networks with pre-processing and post-processing to enable automatic breast MRI segmentation for three tissue types simultaneously: fatty, glandular, and muscular tissue. Pre-processing aims at facilitating training of the network. The architecture of neural network is a Kanezaki-net extended to 3D and consists of two sub-networks. Post-processing is enhancing the obtained segmentations by removing common errors. 25 datasets of T2 weighted MRI from the Medical University of Vienna have been evaluated qualitatively by two observers while eight datasets have been evaluated quantitatively based on a ground truth annotated by a medical practitioner. As a result of the qualitative evaluation, 22 out of 25 are usable for biomechanical models. Quantitatively, we achieved an average dice coefficient of 0.88 for fatty tissue, 0.5 for glandular tissue, and 0.86 for muscular tissue. The proposed method can serve as a robust method for automatic generation of biomechanical models.

Keywords: Breast Magnetic Resonance Imaging, Image Segmentation, Unsupervised Neural Networks, Machine Learning

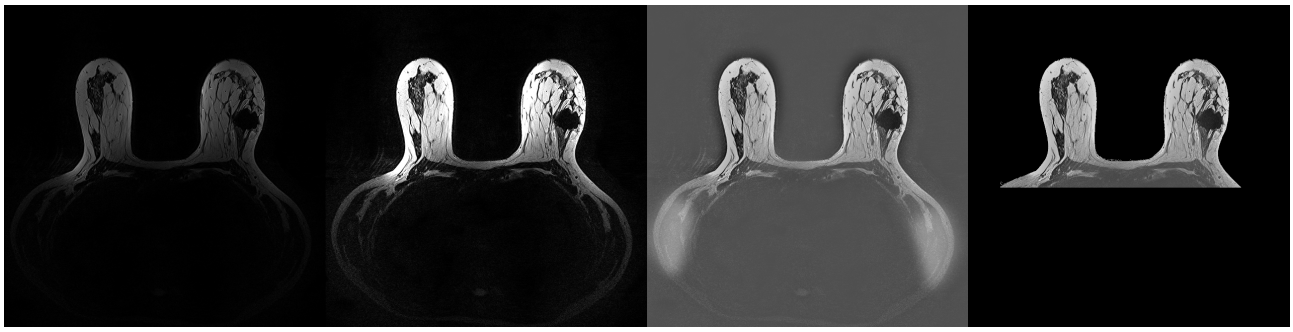
1. INTRODUCTION

For early diagnosis of breast cancer, registration between different modalities such as X-ray mammography, sonography, magnetic resonance imaging (MRI), and spot mammograms has been applied using biomechanical models to cope with the different patient positioning and deformation state of the breast.[1–3] MRI segmentation plays an important role for automatic creation of biomechanical models in such applications.

Several image segmentation techniques for breast MRI have been proposed. One of the techniques is clustering. It could be applied for separating the breast from the background. There are a lot of techniques in clustering such as hierarchical clustering and mean shift clustering. Due to the uncertainty, irregular and fuzzy borders in breast MRI images, K-means is used and Fuzzy C-means (FCM) could be applied to improve segmentation techniques. Besides, gradient and intensity-based clustering could be used in segmentation. However, breast MRI images have noise, intensity inhomogeneity, and weak boundaries which require complex procedures.[4]

In literature, most of the methods proposed for automatic breast segmentation are limited to segmenting two tissue types, fatty and glandular tissue.[5,6] Additionally, there are methods for segmenting muscular tissue.[7] Deep learning algorithms have been recently used for medical image segmentation.[8,9]Apart from the lack of segmentation of muscular tissue, these networks[10,11] need labeled training data for robust functionality which is expensive and time consuming to acquire for 3D breast MRI data.

In this paper, we propose to use an unsupervised neural network (NN) for segmenting three types of tissue simultaneously: fatty, glandular, and muscular tissue. The method was developed for an application in automatic generation of biomechanical models and for segmenting T2-weighted images.



(a) (b) (c) (d)

Figure 1. The steps of pre-processing in one of the slices: The original image from the Medical University of Vienna (a). The enhanced image (b). The corrected bias field image (c). The image multiplied by the K-means mask (d).

2. METHODS

Our proposed methods consist of three parts which are pre-processing, NN classification, and post-processing. First, pre-processing prepares the MR volume for a NN and aims at facilitating training of the network. Second, an iterative approach is used in NNs that generates many different segmentations of the same MR volume. For each iteration, the network is randomly reinitialized and retrained. The NN consists of two sub-networks. Third, post-processing is enhancing the obtained segmentations by removing common errors and selects the best segmentation.

2.1 Pre-processing

Pre-processing consists of four steps: improving contrast, correcting bias field, masking, and shrinking data. The contrast of the images is improved using histogram stretching. It is observed that the intensity distribution in different datasets contains almost 99% of voxels in less than 30% of the intensity range. We assign the remaining 1% of the voxels the maximum intensity of the other 99% of the voxels. Hence, the contrast is improved as shown in Fig. 1 (b). Afterwards the bias field correction based on Fuzzy C-means[12] is used to correct inhomogeneities on the magnetic field as shown in Fig. 1 (c). Since the breast muscle and the background have nearly the same intensity values in T2-weighted MRI, we implemented an air mask using K-means method. Then, an element wise multiplication is done between this mask and the original MRI as seen in Fig. 1 (d).

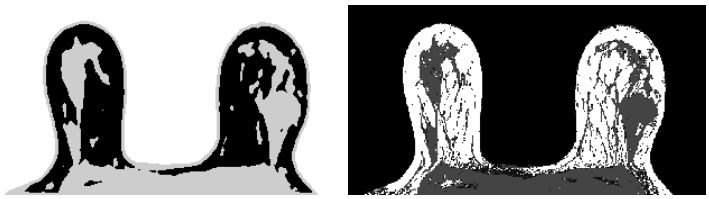
For decreasing the size of the MRI, we are cutting all the voxels of the background from posterior, inferior, left, and right. We rescale the volume of MRI with a factor of 0.623 in order to have less size of memory.

2.2 Neural Networks

Second, our NN architecture is based on the work of Kanezaki et al.[13,14] It is an unsupervised image segmentation method, but has so far only been used for applications outside of medical imaging segmentation and also not specifically for breast MRI. This network consists of convolution layers only, ReLU as an activation function, batch normalization layer and softmax layer. They are cascaded together such that the output of a specific layer is based on the input of the previous layer in a feedforward scheme.

The loss function of this network consists of two criteria. First, pixels with similar features should be assigned to the same class. Cross entropy loss is applied for this purpose. Second, spatially continuous pixels should also be in the same class. The mean absolute error (MAE) is calculated for this purpose between each element in the input x and target y, where x is the difference between the voxel wise feature to its neighbouring voxel and y is a matrix full of zeroes.

We modified this network to be adapted for 3D datasets with anisotropic volume dimensions. An iterative approach is used in NNs that generates different segmentations of the same MR volume because of the random initialization of the NN. For each iteration, the network is randomly reinitialized and retrained and in the



(a) (b)

Figure 2. Segmentation of two sub-networks in one of the slices: The first network of size filter [1x3x3] (a). The second network of size filter [1x1x1] (b).

post-processing the final segmentation is selected from this series of segmentations. We tested empirically that re-initialization ten times with each new input would guarantee general robustness for new data. The loss function of spatial continuity was expanded into the third dimension. Our architecture consists of two sub-networks. The first network with a filter size of [1x3x3] pixels is responsible for classifying the muscular tissue and the second network with a filter size of [1x1x1] pixels is responsible for detecting glandular tissue. The numbers of convolution layers is three and two for the first and the second network, respectively. The learning rate (α) is 0.2 and 0.05 for the first and second network, respectively. The number of channels in both networks is 25. We use a stochastic gradient descent (SGD) optimizer. The number of iterations is 300. We selected these values based on empirical tests on our available data.

2.3 Post-processing

Third, post-processing is enhancing the obtained segmentations by removing common errors and selecting the best segmentation. Since the learning is performed unsupervised, there is no specific class label for each tissue. We make use of heuristic observations to sort the segmented areas by the NN to the tissue labels: 0 (background), 1 (glandular), 2 (fatty), and 3 (muscle). Since the background contains the majority of the voxels, we assign the segmented area with the largest number of voxels to label 0. Due to the characteristics of T2 weighted MRI, the fatty tissue is the brightest tissue. Hence we analyze the 100 brightest pixels in the original MRI dataset and assign the NN tissue class of which these pixels are members to label 2.

Muscular and glandular tissue are often assigned to the same class due to their similar intensity value in T2 weighted images. Hence, we extract a binary mask of the mixed class of glandular and muscular tissues from the sub-network of filter size [1x3x3]. It can be empirically seen in all our available datasets that the breast muscle could be treated as one single object of a large area while the glandular tissue consists of multiple small areas. Hence in each slice, we obtain the largest object and assign it to label 3. The remaining objects are assigned to label 1.

For selecting the best segmentation from the ten iterations of running the sub-networks, we propose a selection metric to judge the quality of the muscular and glandular tissue segmentation. We calculate a muscle and glandular quality score by a combination of several heuristic metrics: for muscular tissue, we first evaluate the muscle occurrence in each segmentation. Second, we evaluate the symmetry of the muscle. Third, we calculate the number of erosion steps that were done for the inaccurate detection of the muscular tissue in the anterior part of the breast.

For glandular tissue, we first evaluate the glandular occurrence in each segmentation. Second, we calculate the overlap of the muscular tissue between the two sub-networks. Third, for the position of the 100 brightest pixels, we calculate a ratio of those pixels between the default labels (fatty tissue) and the actual segmented labels. Fourth, we calculate the percentage of glandular tissue at the borders of the breast and in the area below the sternum position. We weighted these values of these metrics for both muscular and glandular tissue based on empirical tests on our available data.

(a) (b) (c) (d)

Figure 3. The steps of post-processing in one of the slices: The input for the post-processing block from NNs [1x3x3] (a). The assigned classes to each tissue(b). Separating muscle to be in a different class (c). The glandular tissue only (d).

3. RESULTS

We tested our methods qualitatively using 25 datasets by two observers and quantitatively using eight datasets of T2-weighted MRI images of different resolution from the Medical University of Vienna. Datasets have been acquired on a total of ten different models of Siemens MRI devices.

For qualitative results, we split slices equally in each dataset into three categories: inferior, middle, and superior slices. We investigated seven problems that could appear in the tissues in each category. For muscular tissue, there were three problems: differentiating between glandular and muscular tissue, holes in the muscle, and artifacts at the border between glandular and muscular tissue. For glandular tissue, there were two problems: noisy classification, and the borders of the breast could be falsely detected as glandular tissue due to the partial volume effect. For fatty tissue, there were two problems: a wrong overall shape of the breast and missing parts of the breast tissue. We estimated a score from 1 to 5 (worst to best) in each problem and took the average of two observers together. We summed over the scores for all seven considered problems resulting in a maximum score of 35 and a minimum score of seven. Then, we categorized the total score of each category into five groups as shown in Table 1. It is realized that more problems are in inferior and superior slices while fewer problems are in middle slices. In category 30-35, more than 75% of the middle slices are without major problems. In category 15-20, only 12% and 8% of inferior and superior slices, respectively are with problems in segmentation.

For quantitative results, we compared our final 3D segmentations to a ground truth segmentation. The ground truth was created by first using niftySeg[5] for classifying glandular and fatty tissue. Then it was corrected manually by a medical practitioner using MITK[15] and the breast muscle was manually segmented in each slice. We evaluated the full segmentation using the DICE Similarity Coefficient (DSC).[16] The mean of DSC is 0.88 for fatty, 0.5 for glandular, and 0.86 for muscular tissue.

To validate numerically the three categories from our qualitative results, we calculate the mean DSC of eight clinical datasets for the three tissues in the inferior, middle, and superior slices as shown in Fig. 4. It shows that the mean DSC in the middle slices are the most robust category which confirms the evaluation by the two observers.

We compared our methods to other segmentation approaches by replacing only the 3D NNs with common algorithms: Fuzzy C-means,[1] and K-means as shown in Fig. 5. In all of the next comparisons, the result of those algorithms are passing our pre-processing and post-processing. For Fuzzy C-means with three clusters, the mean DSCs for the three tissues are lower than our results based in 3D NNs while the results for four clusters are even worse than with three clusters. K-means with three clusters shows less accuracy in all metrics than our method while with four clusters, for fatty and glandular tissue, it is comparable to our methods but muscle

Table 1. Percentage of datasets in the group score of each category of slices

Slices	**30-35**	**25-30**	**20-25**	**15-20**	**0-15**
Inferior	56%	32%	-	12%	-
Middle	76%	12%	12%	-	-
Superior	64%	16%	12%	8%	-

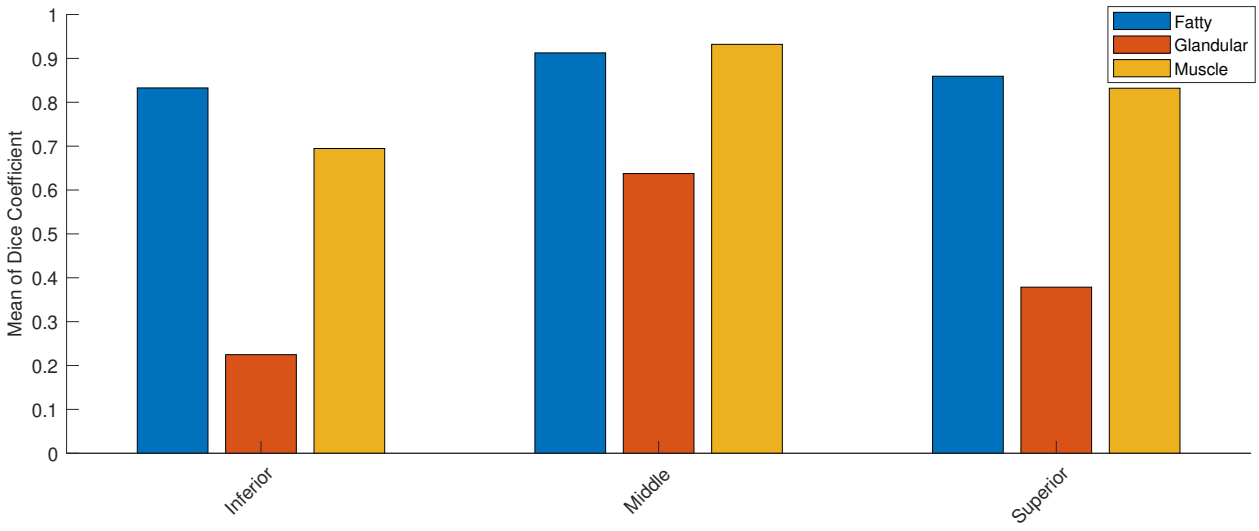


Figure 4. Mean DSC of eight clinical datasets for the three tissues: fatty, glandular, and muscular tissues in inferior, middle, and superior slices, respectively.

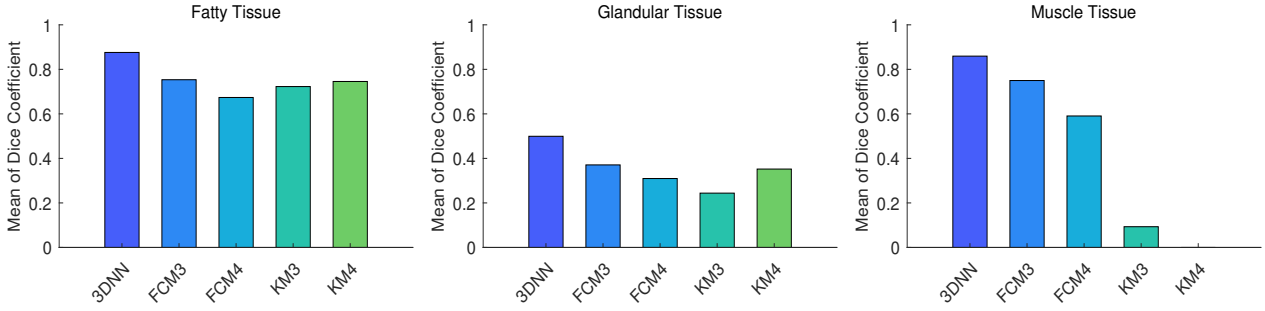


Figure 5. Mean DSC of each tissue (fatty, glandular, and muscle) for six datasets in our proposed algorithm (3DNN) compared to Fuzzy C-means three clusters (FCM3) and four clusters (FCM4) and K-means three clusters (KM3) and four clusters (KM4).

is not detected at all. An exemplary segmentation compared to Fuzzy C-means with three clusters is shown in Fig. 6.

While evaluated on different databases, we compared our method to algorithms from literature as shown in Table 2 to give a relation to other approaches. All of them are restricted to one or two tissues. The first method was to estimate breast density for two tissues only.[6] One method used to detect the muscular tissue using atlas-based breast muscle segmentation method.[7] Two of these methods are using supervised learning which needs quite a lot of labeled data for training.[10,11]

An example of the deformed MRI from our biomechanical model using the segmentation of NNs compared

Table 2. Comparison of DSC for three tissues compared to literature review

Tissue	Gubern-Mérida, et al.[6]	Gubern-Mérida, et al.[7]	Zhang, et al.[10]	Ha, et al.[11]	**our**
Fatty	0.94	-	0.86	-	**0.88**
Glandular	0.80	-	0.83	0.81	**0.50**
Muscular	-	0.72-0.74	-	-	**0.86**

(a) (b) (c)

Figure 6. The ground truth from the medical practitioner (a). The segmentation of our proposed methods (b). The segmentation of Fuzzy C-means with three clusters (c).

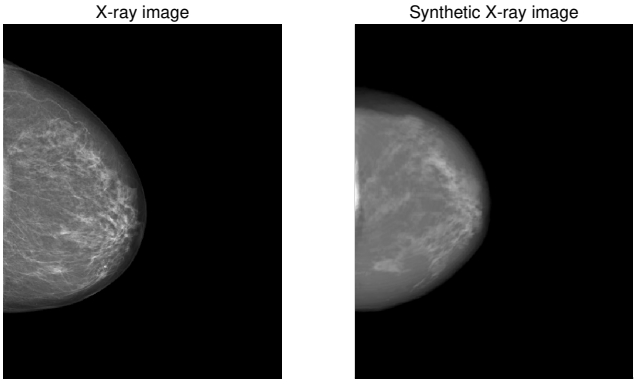


Figure 7. An example showing how segmentation using NNs will help in creating the deformed MRI image from our biomechanical model: X-ray image (left) and Synthetic X-ray (deformed MRI) (right)

to the original X-ray image is shown in Fig. 7.

4. DISCUSSION AND CONCLUSION

The proposed method provides a novel automatic approach using unsupervised NNs for the segmentation of three tissue types in T2 weighted breast MRI concurrently. The method shows very promising results with a mean DSC of 0.88, 0.5, and 0.86 for fatty, glandular, and muscular tissue, respectively. While fatty and muscular tissue segmentation performance is similar to literature values, the mean DSC for glandular tissue is comparably lower. We think that this is partly an effect of statistics, since due to the lower number of voxels in this class, errors are overemphasized. Moreover the ground truth segmentation does not capture tiny details of glandular structures, which however are partly segmented with our method as shown in Fig. 6 and thereby may lead to a reduced DSC score while actually obtaining very good results. The agreement between glandular structures obtained from our segmentation is showcased in Fig. 7 in comparison to the X-ray mammogram of a patient. As a survey of qualitative results of 25 datasets, it shows very promising scores. In only three out of 25 datasets, it failed to produce a usable result for automatic generation of a biomechanical model. In future, we would like to extend our database of labeled datasets for more quantitative validation of the method. The method may also be extended to different MRI series. Until now a segmentation can be computed in range of few minutes for one dataset, though the emphasis of this paper was not on computation time. Using unsupervised learning will allow us to segment breast MRI without the need of having labeled data for training. It therefore serves as a fundamental robust method for automatic generation of patient-specific biomechanical models.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) under grant number 5565/3-1 and the Austrian Science Fund (FWF) under grant number 4240.

REFERENCES

[1] Hopp, T., Dietzel, M., Baltzer, P., Kreisel, P., Kaiser, W., Gemmeke, H., and Ruiter, N., "Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity-based optimization," *Medical Image Analysis* **17**(2), 209–218 (2013).

[2] Hopp, T., Baltzer, P., Dietzel, M., Kaiser, W., and Ruiter, N., "2D/3D image fusion of X-ray mammograms with breast MRI: Visualizing dynamic contrast enhancement in mammograms," *International Journal of Computer Assisted Radiology and surgery* **7**, 339–48 (06 2011).

[3] Said, S., Clauser, P., Ruiter, N. V., Baltzer, P. A. T., and Hopp, T., "Image registration between MRI and spot mammograms for X-ray guided stereotactic breast biopsy: preliminary results," in [*Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*], **11598**, 354 – 361, SPIE (2021).

[4] Despotovic, I., Goossens, B., and Philips, W., "MRI segmentation of the human brain: Challenges, methods, and applications," *Computational and Mathematical Methods in Medicine* **2015** (2015).

[5] Cardoso, M., Clarkson, M., Modat, M., and Ourselin, S., "Niftyseg: Open-source Software for Medical Image Segmentation, Label Fusion and Cortical Thickness Estimation," in [*IEEE International Symposium on Biomedical Imaging, Barcelona, Spain*], (2012).

[6] Gubern-Mérida, A., Kallenberg, M., Mann, R. M., Martí, R., and Karssemeijer, N., "Breast Segmentation and Density Estimation in Breast MRI: A Fully Automatic Framework," *IEEE Journal of Biomedical and Health Informatics* **19**(1), 349–357 (2015).

[7] Gubern-Mérida, A., Kallenberg, M., Martí, R., and Karssemeijer, N., "Segmentation of the Pectoral Muscle in Breast MRI Using Atlas-Based Approaches," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*], Ayache, N., Delingette, H., Golland, P., and Mori, K., eds. (2012).

[8] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., and Nandi, A. K., "Medical Image Segmentation Using Deep Learning: A Survey," *IET Image Processing* (Jan 2022).

[9] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., and Terzopoulos, D., "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 1–1 (2021).

[10] Zhang, Y., Chen, J.-H., Chang, K.-T., Park, V. Y., Kim, M. J., Chan, S., Chang, P., Chow, D., Luk, A., Kwong, T., and Su, M.-Y., "Automatic Breast and Fibroglandular Tissue Segmentation in Breast MRI Using Deep Learning by a Fully-Convolutional Residual Neural Network U-Net," *Academic Radiology* **26**(11), 1526–1535 (2019).

[11] Ha, R., Chang, P., Mema, E., Mutasa, S., Karcich, J., Wynn, R. T., Liu, M. Z., and Jambawalikar, S., "Fully Automated Convolutional Neural Network Method for Quantification of Breast MRI Fibroglandular Tissue and Background Parenchymal Enhancement," *Journal of Digital Imaging* **32**, 141–147 (Feb 2019).

[12] Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., and Moriarty, T., "A modified Fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging* **21**(3), 193–199 (2002).

[13] Kim, W., Kanezaki, A., and Tanaka, M., "Unsupervised Learning of Image Segmentation Based on Differentiable Feature Clustering," *IEEE Transactions on Image Processing* **29**, 8055–8068 (2020).

[14] Kanezaki, A., "Unsupervised image segmentation by backpropagation," in [*2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 1543–1547 (2018).

[15] Nolden, M., Zelzer, S., Seitel, A., Wald, D., Müller, M., Franz, A. M., Maleike, D., Fangerau, M., Baumhauer, M., Maier-Hein, L., Maier-Hein, K. H., Meinzer, H.-P., and Wolf, I., "The Medical Imaging Interaction Toolkit: challenges and advances: 10 years of open-source development," *International Journal of Computer Assisted Radiology and Surgery* **8**, 607—620 (July 2013).

[16] Zou, K., Warfield, S., Bharatha, A., Tempany, C., Kaus, M., Haker, S., Wells, W., Jolesz, F., and Kikinis, R., "Statistical validation of image segmentation quality based on a spatial overlap index.," *Academic Radiology* **11 2**, 178–89 (2004).