Image Synthesis-based Late Stage Cancer Augmentation and Semi-Supervised Segmentation for MRI Rectal Cancer Staging

Saeko Sasuga[1], Akira Kudo[1], Yoshiro Kitamura[1], Satoshi Iizuka[2], Edgar Simo-Serra[3], Atsushi Hamabe[4], Masayuki Ishii[4], and Ichiro Takemasa[4]

[1] Imaging technology center, Fujifilm corporation, Tokyo, Japan
[2] Center for Artificial Intelligence Research, University of Tsukuba, Ibaraki, Japan
[3] Department of Computer Science and Engineering, Waseda University, Tokyo, Japan
[4] Department of Surgery, Surgical Oncology and Science, Sapporo Medical University, Hokkaido, Japan
saeko.sasuga@fujifilm.com

Abstract. Rectal cancer is one of the most common diseases and a major cause of mortality. For deciding rectal cancer treatment plans, T-staging is important. However, evaluating the index from preoperative MRI images requires high radiologists' skill and experience. Therefore, the aim of this study is to segment the mesorectum, rectum, and rectal cancer region so that the system can predict T-stage from segmentation results.

Generally, shortage of large and diverse dataset and high quality annotation are known to be the bottlenecks in computer aided diagnostics development. Regarding rectal cancer, advanced cancer images are very rare, and per-pixel annotation requires high radiologists' skill and time. Therefore, it is not feasible to collect comprehensive disease patterns in a training dataset. To tackle this, we propose two kinds of approaches of image synthesis-based late stage cancer augmentation and semi-supervised learning which is designed for T-stage prediction. In the image synthesis data augmentation approach, we generated advanced cancer images from labels. The real cancer labels were deformed to resemble advanced cancer labels by artificial cancer progress simulation. Next, we introduce a T-staging loss which enables us to train segmentation models from per-image T-stage labels. The loss works to keep inclusion/invasion relationships between rectum and cancer region consistent to the ground truth T-stage. The verification tests show that the proposed method obtains the best sensitivity (0.76) and specificity (0.80) in distinguishing between over T3 stage and underT2. In the ablation studies, our semi-supervised learning approach with the T-staging loss improved specificity by 0.13. Adding the image synthesis-based data augmentation improved the DICE score of invasion cancer area by 0.08 from baseline. We expect that this rectal cancer staging AI can help doctors to diagnose cancer staging accurately.

Keywords: Rectal cancer · Segmentation · T stage discrimination · Semi-supervised learning · Image synthesis.

1   Introduction

T-staging is an index for assessing the spread of primary tumors into nearby tissues. Regarding rectal cancer which is a common disease and a major cause of mortality. Magnetic resonance imaging (MRI) is commonly used for the assessment of T2/T3/T4 staging. In the clinical research [3, 16] , it has become clear that invasion depth from rectum and margin between cancer and mesorectum are directly related to the prognosis of rectal cancer, and those indexes are also important for deciding rectal cancer treatment plan. However, evaluating T-staging from preoperative MRI images requires high radiologists' skill and experience. In [13], the T3 staging sensitivity and specificity for experienced gastrointestinal radiologists is 96% and 74%, while for a general radiologist is 75% and 46%, respectively. Since those numbers are not sufficiently high, computer aided diagnosis (CAD) systems for assisting T-staging were proposed in the previous studies [10].

This paper deals with two difficulties in building such CAD systems. First one is that preparing per-pixel annotation requires high radiologists' skill and time cost. Second one is that advanced cancer cases are very rare. Due to these difficulties, it is almost impossible to collect large scale high quality training dataset. To tackle these problems, we propose semi-supervised learning and image synthesis-based late stage cancer augmentation. Here we pick up related works, then summarize contributions of this work.

Cancer Segmentation and T-staging Since the U-net [14] was proposed, segmentation tasks have achieved remarkable results. Regarding tumor segmentation tasks from MR images, it is reported that segmentation accuracy reached human expert level [5, 7, 9]. However, not so many researches have been done on automatic T-staging. One research [4] tried to segment tumor, inner wall, and outer wall to visualize cancer invasion, but it did not evaluate T-staging accuracy. Kim et al. [10] proposed a method for classifying T-stage from preoperative rectal cancer MRI, but their method was not able to visualize the basis of the decision. To the best of our knowledge, there was no previous study that simultaneously evaluate both cancer segmentation and T-staging accuracy.

Data augmentation Supervised learning methods rely on a large and diverse training dataset to achieve high performance. Data augmentation is commonly known as an effective way to compensate limited amount of training data [15].Recently, generation-based approach using generative adversarial networks (GANs) [6] have been proposed. One such method is label-to-image translation. Abhishek et al. [1] used label-to-image translation method to generate skin cancer 2D images for data augmentation and showed improvement in segmentation accuracy. However, this method has limitations that the generated data follow training data distribution. It is difficult to create images that are rarely included in training data.
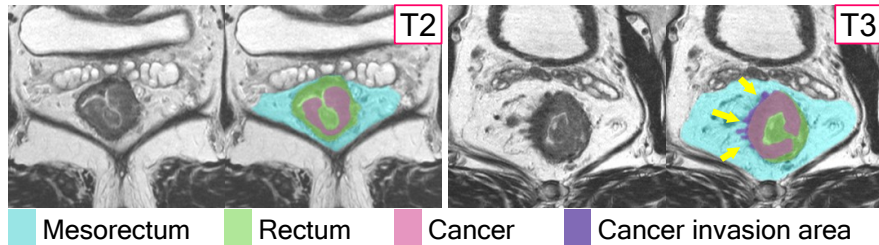We summarize the contributions of this work as follows:

Mesorectum   Rectum   Cancer   Cancer invasion area

Fig. 1. T-stage classification algorithm using multi-label relationship. Yellow arrows point cancer area extends beyond the rectum area and this case is staged with T3.

1. This is an early work which develops and evaluates a CAD system that can do both rectal cancer segmentation and T-staging with a clinical dataset.
2. We propose a semi-supervised learning using a novel T-staging loss. The T-staging loss enables us to train a segmentation model with not only per-pixel labels but per-image T-stage labels.
3. We propose a label-to-image translation based severe cancer augmentation. This approach can generate rare data that is out of training dataset distribution by deforming cancer label into the shape of a progressive cancer.

2   Methodology

We model the rectal cancer T-staging problem as a multi-class multi-label segmentation of mesorectum, rectum, and rectal cancer as shown in Figure 1. The algorithm predicts T-stage following to the segmentation results. When the cancer area is not in contact with the contour of the rectum area, the case is classified as under T2 stage. On the other hand, when the cancer area crosses over the rectum area, the case is classified as over T3 stage. This rule exactly follows the T-staging rules of tumor invasion into the area of the rectum. In this section, we explain a semi-supervised approach with a novel loss function named T-staging loss in Section 2.1. Then, we explain image synthesis-based late stage cancer augmentation in Section 2.2.

2.1   Semi Supervised Learning with T-staging Loss

The proposed semi supervised learning architecture is shown in Figure 2. We use a 3D variant of U-net that we train to perform a segmentation task. The last three channels of the network are the probabilities of mesorectum, rectum and rectal cancer areas. T-staging results are calculated based on the binarized segmentation results. In the training phase, 3D MR images with ground-truth segmentation labels or just only ground truth T-stage are inputted to the network. The loss function consists of a segmentation loss and the proposed T-staging loss; $Loss_{SEG} + \lambda \times Loss_{STG}$, where λ is a parameter used to balance the two
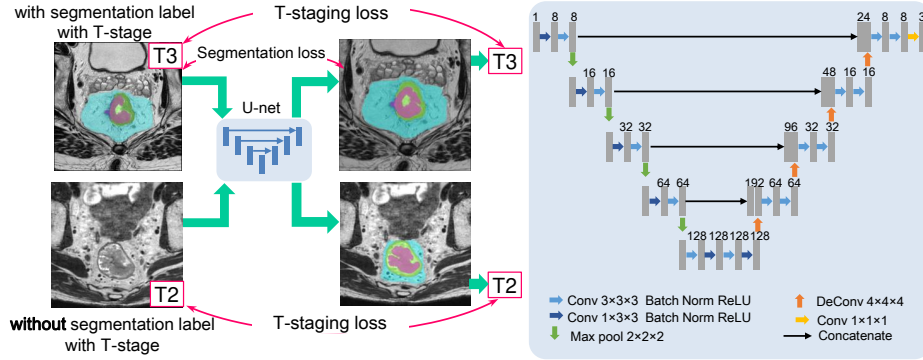
Fig. 2. Frameworks of our semi-supervised learning and the architecture of segmentation network. In the training phase, images with and without segmentation labels are inputted to the U-net. Training data with only T-stage ground truth are used for calculating T-staging loss which works as to maximize correspondence between ground truth and prediction T-stage.

terms. We used a standard Dice loss [11] for $Loss_{SEG}$. The T-staging loss which for accurate staging purposes is defined as follows:

$$Loss_{STG} = -\mu \log \frac{p_{STG}{}^{g_{STG}}}{(1 - p_{STG})^{(g_{STG}-1)}} + \frac{\sum_{i=1}^{N} p_{invT2i} + \alpha}{\sum_{i=1}^{N} p_{invT2i} + \sum_{i=1}^{N} p_{invT3i} + \alpha}. \quad (1)$$

Where,

$$p_{STG} = \max(p_{cancer,i} \times (1 - p_{rectum,i})). \quad (2)$$

$$g_{STG} = \begin{cases} 1 & \text{if ground truth T stage is over T3} \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

$$p_{invT2} = p_{cancer} \times (1 - p_{rectum}) \times (1 - g_{staging}). \quad (4)$$

$$p_{invT3} = p_{cancer} \times (1 - p_{rectum}) \times (1 + g_{staging}). \quad (5)$$

N is the number of voxels. p_{cancer} and p_{rectum} represent the probability maps of the rectal cancer and rectum, respectively. $p_{staging}$ indicates the probability of the predicted staging. It takes a high value when there is any voxel simultaneously having low rectum probability and high cancer probability. p_{invT2} and p_{invT3} are probabilities of cancer invasion of each voxel in T2 case and T3 case. This term works to reduce the cancer area outside of the rectum for T2 cases. On the other hand, it works to increase the cancer area outside of the rectum for T3 cases. In contrast to $Loss_{SEG}$ is used only for the images with ground-truth segmentation labels. $Loss_{STG}$ can be used for all cases having only ground truth T-staging. So, it works as semi-supervised setting in segmentation.
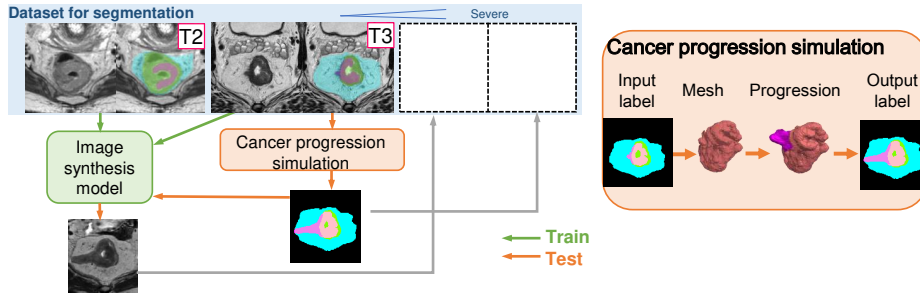
Fig. 3. Overview of severe cancer data syntheses framework. The image synthesis model is trained with real images. On the test phase, cancer labels are modified by simulation. Following the generated labels, image synthesis model generate severe cancer images.

2.2   Generating Advanced Cancer MRI Image from Labels

Figure 3 is an overview of advanced cancer image synthesis system. In general, GAN can model the underlying distribution of training data. To the contrary, our goal is to generate out-of-distribution data (e.g., we have relatively many early stage cancer, but we want more advanced staged cancer images). Therefore, we first train a label-conditional-GAN model. Then we generate images for data augmentation by inputting various deformed label images. An important point is that the deformation is done by external knowledge which simulate cancer development.

Semantic image synthesis SPADE [12] is known as one of the best architectures for the semantic image synthesis task. However, it is rarely applied on 3D images due to their large computational cost. This is an early work applying SPADE for 3D image synthesis. Overall, we adopt and extended the basic framework of SPADE from 2D to 3D. As shown in Figure 4, we simply introduce a SPADE-3D block in each scale of the generator instead of residual-SPADE block which was proposed in the original paper to save computational cost.

We found that the more the kinds of input label classes, the more generated images give realistic looking and anatomically consistent. For that reason, in addition to cancer, we also used rectum, mesorectum, bladder, prostate and pelvis classes as the input to the generator. We monitored the training until the generator gives realistically looking images from the given test label images.

Cancer label deformation Advanced stage cancer labels are generated as follows. First, a cancer label was transformed to a 3D mesh model consisting of vertices and faces. Next, the furthest invasion point, infiltration direction vector, and non-deformable part were calculated based on cancer morphology [8]. Then, the 3D mesh model was transformed by constraint shape optimization [2]. Finally, the mesh model was voxelized again. These processes were applied several times for each image, randomly changing a vertex to move and shifting distance.
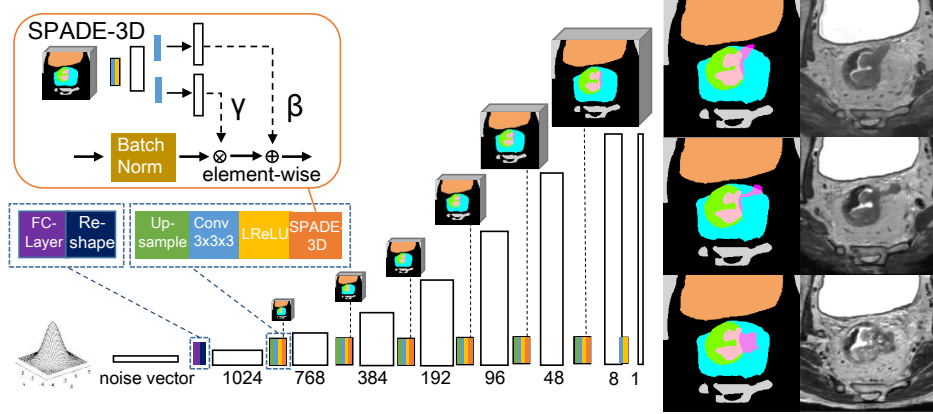
Fig. 4. Overview of the 3D generator architecture. Each normalization layer uses 3D label image to modulate the layer activations. To save computation cost, we removed residual blocks in the original SPADE [12] and introduced a single SPADE-3D block per scale which had enough capacity to generate realistic images.

Fig. 5. Samples of the input labels and generated images. Different images are generated based on the deformed cancer labels.

We also mimicked the shape of the cancer infiltrate into vessels or lymph nodes by randomly embedded the tubular or spherical objects to the cancer model. The MR images corresponding to these labels were variously generated by the GAN as shown in Figure 5 and supplementary movies.

3 Experiments and Results

3.1 Dataset

We curated four types of datasets A-D. The dataset A-C were real MR images acquired from rectal cancer patients. The dataset D was generated by a GAN model which was trained on dataset A. The dataset B was separated from A for a fair comparison, meaning that the dataset D was not generated from testing dataset B. The dataset A, B, and D had per-pixel ground truth labels, and dataset C had only T-stage label. The number of samples in each dataset are provided in Table 1.

MR T2-weighted images were acquired using a 3.0 T (N=89) or 1.5 T (N=106) MR scanner. Ground truth segmentation labels were annotated by two surgeons based on the pathological specimens removed by surgeries. Ground truth T-stage labels were labeled based on the pathological diagnosis.

3.2 Implementation Details and Evaluation Metrics

During the training, the Adam optimizer with a learning rate of 0.003 was used. The parameters in the loss function are as follows; λ=0.1, μ=0.1, α=500. Mini

Table 1. The dataset specification and usage in this study.

	Patient Num. (T2/T3)	Data type Image type	Ground truth	Usage Segmentation	Image Synthesis
A	69 (22/47)	Real	Label	Train	Train
B	66 (26/40)	Real	Label	Train/Evaluation	Unused
C	60 (29/31)	Real	T-Stage	Train/Evaluation	Unused
D	78 (0/78)	Generated	Label	Train	-

Table 2. The condition of ablation studies.

	Loss function	Dataset
Baseline (DICE loss)	Soft dice	A,B
+Semi supervised	Soft dice+T-staging loss	A,B,C
+Data augmentation	Soft dice+T-staging loss	A,B,C,D

batch consisted of 4 cases with segmentation labels and 2 cases with only T-stage. Experiments were conducted up to 80000 iterations.

For the evaluation, five-fold cross validation was conducted on dataset B and C. Those evaluation datasets were randomly divided into five subgroups. Four subgroups out of the five and dataset A, and D were used for training the segmentation network. Note that one subgroup in the four subgroups for training was used for validation. We chose the best model according to the DICE score of the validation subgroup, then applied the model to the remaining subgroup. We used two indexes of Dice similarity coefficient in volume [17], and T-staging sensitivity/specificity for evaluation. We calculated T-staging sensitivity as a correctly predicted T3 number divided by the ground truth T3 number. The specificity is a correctly predicted T2 number divided by the ground truth T2 number.
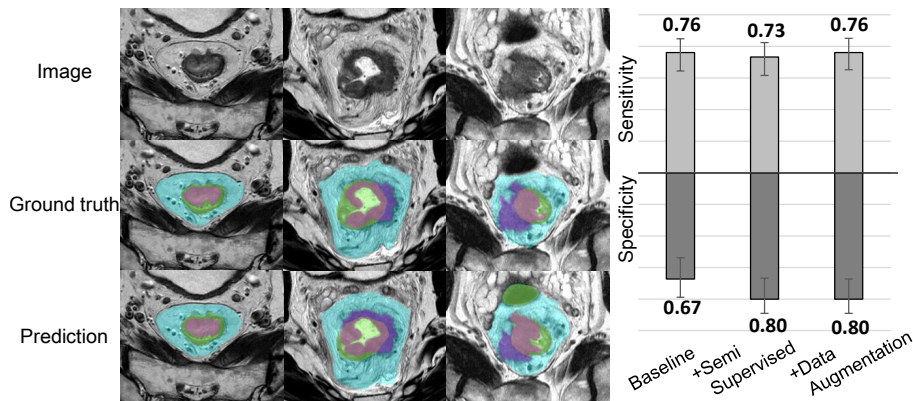
3.3   Results

Ablation studies were conducted to evaluate the efficacies of the proposed methods. We compared three experimental settings shown in Table 2. Each setting corresponds to the baseline (the DICE loss [11]), baseline plus semi-supervised approach, and baseline plus semi-supervised approach and late stage data augmentation.

We compared the three settings with the Dice coefficients of mesorectum, rectum, cancer and cancer invasion area. Table 3 shows the results of dice coefficient. Figure 6 shows several segmentation results. Both of the semi-supervised and the data augmentation increased the DICE score of cancer area. Especially, they were effective in cancer invasion area. The segmentation results of mesorectum and rectum were not affected by the proposed approach, as is expected. Figure 7 shows results of T-staging accuracy. The proposed method obtains the best sensitivity (0.76) and specificity (0.80), and improved specificity by a margin of 0.13.

Table 3. Comparison of DICE score in ablation study.

	Mesorectum	Rectum	Cancer	Invasion area
Baseline (DICE loss)	0.907	0.908	0.642	0.415 [0.330-0.501]
+Semi supervised	**0.910**	0.910	0.669	0.449 [0.380-0.519]
+Data augmentation	0.903	**0.911**	**0.680**	**0.495 [0.413-0.577]**

Fig. 6. Example segmentation results. The left column shows T2 cases, the center column is T3 case and the right column is a severe cancer case.

Fig. 7. Comparison of T-staging sensitivity and specificityin ablation study.

4   Discussion

We assumed that the number of available per-pixel training data is small. This often happens at clinical practice. In fact, since rectal cancer is ambiguous on MR images, the surgeons annotated labels with reference to pathological specimens to make reliable ground truth in this research. The T-staging accuracy of the proposed methods was not high enough compared to pathologically proved ground truth. One reason is that T-staging is a hard task for a radiologist too. Another one is that the proposed data augmentation handles only rare shape of the cancer. The future work is generating not only rare shape but rare texture. This study was conducted on data acquired from one site. We are going to validate the algorithm with multi-site data and severe real cancer data.

5   Conclusions

In this paper, we proposed a novel semi-supervised learning method designed for cancer T-staging, and image synthesis-based data augmentation to generate advanced cancer images. Ablation studies showed that the methods improved rectal cancer segmentation and cancer T-staging accuracy. We expect that this cancer segmentation would help doctors diagnose T-staging in clinical practice.

References

1. Abhishek, K., Hamarneh, G.: Mask2lesion: Mask-constrained adversarial skin lesion image synthesis. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 71–80. Springer (2019)
2. Botsch, M., Kobbelt, L.: An intuitive framework for real-time freeform modeling. ACM Transactions on Graphics (TOG) **23**(3), 630–634 (2004)
3. Cho, S.H., Kim, S.H., Bae, J.H., Jang, Y.J., Kim, H.J., Lee, D., Park, J.S.: Prognostic stratification by extramural depth of tumor invasion of primary rectal cancer based on the radiological society of north america proposal. American Journal of Roentgenology **202**(6), 1238–1244 (2014)
4. Dolz, J., Xu, X., Rony, J., Yuan, J., Liu, Y., Granger, E., Desrosiers, C., Zhang, X., Ben Ayed, I., Lu, H.: Multiregion segmentation of bladder cancer structures in mri with progressive dilated convolutional networks. Medical physics **45**(12), 5482–5493 (2018)
5. Feng, X., Tustison, N.J., Patel, S.H., Meyer, C.H.: Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. Frontiers in computational neuroscience **14**, 25 (2020)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
7. Hodneland, E., Dybvik, J.A., Wagner-Larsen, K.S., Šoltészová, V., Munthe-Kaas, A.Z., Fasmer, K.E., Krakstad, C., Lundervold, A., Lundervold, A.S., Salvesen, Ø., et al.: Automated segmentation of endometrial cancer on mr images using deep learning. Scientific reports **11**(1), 1–8 (2021)
8. Horvat, N., Carlos Tavares Rocha, C., Clemente Oliveira, B., Petkovska, I., Gollub, M.J.: Mri of rectal cancer: Tumor staging, imaging techniques, and management. RadioGraphics **39**(2), 367–387 (2019). https://doi.org/10.1148/rg.2019180114, https://doi.org/10.1148/rg.2019180114, pMID: 30768361
9. Huang, Y.J., Dou, Q., Wang, Z.X., Liu, L.Z., Jin, Y., Li, C.F., Wang, L., Chen, H., Xu, R.H.: 3-d roi-aware u-net for accurate and efficient colorectal tumor segmentation. IEEE Transactions on Cybernetics **51**(11), 5397–5408 (2020)
10. Kim, J., Oh, J.E., Lee, J., Kim, M.J., Hur, B.Y., Sohn, D.K., Lee, B.: Rectal cancer: Toward fully automatic discrimination of t2 and t3 rectal cancers using deep convolutional neural network. International Journal of Imaging Systems and Technology **29**(3), 247–259 (2019)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
12. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2337–2346 (2019)
13. Rafaelsen, S.R., Sørensen, T., Jakobsen, A., Bisgaard, C., Lindebjerg, J.: Transrectal ultrasonography and magnetic resonance imaging in the staging of rectal cancer. effect of experience. Scandinavian journal of gastroenterology **43**(4), 440–446 (2008)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

15. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of big data **6**(1), 1–48 (2019)
16. Taylor, F.G., Quirke, P., Heald, R.J., Moran, B.J., Blomqvist, L., Swift, I.R., Sebag-Montefiore, D., Tekkis, P., Brown, G.: Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the mercury study. Journal of Clinical Oncology **32**(1), 34–43 (2014)
17. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells III, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Academic radiology **11**(2), 178–189 (2004)