

PROSTATE CANCER DETECTION AND SEGMENTATION IN MULTI-PARAMETRIC MRI VIA CNN AND CONDITIONAL RANDOM FIELD

Ruiming Cao^{*†} Xinran Zhong^{*} Sepideh Shakeri^{*} Amirhossein Mohammadian Bajgiran^{*}
 Sohrab Afshari Mirak^{*} Dieter Enzmann^{*} Steven S. Raman^{*} Kyunghyun Sung^{*}

^{*} Radiology, University of California - Los Angeles, Los Angeles, CA 90095, USA

[†] Computer Science, University of California - Los Angeles, Los Angeles, CA 90095, USA

ABSTRACT

Multi-parametric MRI (mp-MRI) is a powerful diagnostic tool for prostate cancer (PCa). However, interpreting prostate mp-MRI requires high-level expertise, causing significant inter-reader variations. Convolutional neural networks (CNNs) have recently shown great promise for various tasks. In this study, we propose an improved CNN to jointly detect PCa lesions and segment for accurate lesions contours. Specifically, we adapt focal loss to overcome the imbalance between cancerous and non-cancerous areas for improved lesion detection and design selective dense conditional random field (SD-CRF), a post-processing step to refine the CNN prediction into the lesion segmentation based on a specific imaging component of mp-MRI. We trained and validated the proposed CNN in 5-fold cross-validation using 397 pre-operative mp-MRI exams with whole-mount histopathology-confirmed lesion annotations. In the free-response receiver operating characteristics (FROC) analysis, the proposed CNN achieved 75.1% lesion detection sensitivity at the cost of 1 false positive per patient. In the evaluation for lesion segmentation, the proposed CNN improved the Dice coefficient by 20.6% from the baseline CNN.

Index Terms— Computer-aided detection and diagnosis, prostate cancer, MRI, convolutional neural networks

1. INTRODUCTION

Prostate cancer (PCa) is the most diagnosed cancer among men in the United States [1]. Multi-parametric MRI (mp-MRI) provides a powerful combination of anatomical and functional information and serves as a non-invasive imaging tool for the diagnosis of PCa. Computerized analysis of mp-MRI is an active research area to complement and overcome limitations of the current qualitative interpretation of mp-MRI [2, 3]. Tsehay *et al.* [4] and Kiraly *et al.* [5] demonstrated the improved performance over imaging feature-based

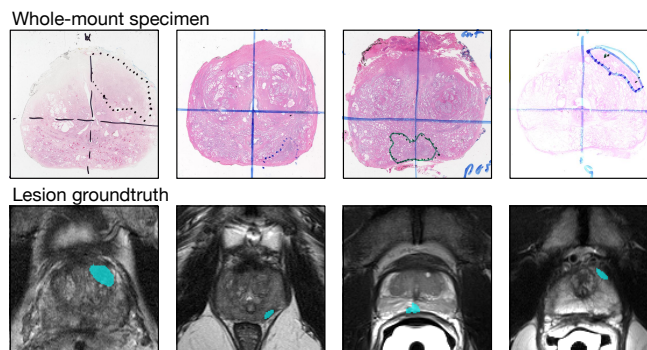


Fig. 1. The lesion groundtruth was retrospectively annotated using whole-mount histopathology as the reference.

methods using convolutional neural networks (CNNs), but the tasks were focused on detection and classification only. Although it is highly desirable to jointly detect and segment the PCa lesions since the location, size, and shape of the lesion play important roles in the diagnosis and treatment planning of PCa [6], the methods that can jointly detect and segment PCa lesions were not well investigated, mainly due to challenges such as the highly imbalanced number of normal and cancerous voxels and difficulties associated with different intensity and contrast patterns in the multi-parametric imaging.

Here, we describe the novel CNN model to jointly detect and segment PCa lesions, overcoming the challenges. Firstly, when the number of normal voxels is much higher than the number of cancerous voxels (e.g., only 1.6% of all voxels are annotated as cancerous in our study), adequate training of CNNs becomes difficult because the total loss is mostly composed of normal voxels. We use focal loss (FL), a modified cross-entropy loss, to adaptively controls the weights for each voxel [7] and thus achieve balanced training from both normal and cancerous voxels. Secondly, as CNNs predict the probability map using multiple imaging components of mp-MRI, when a lesion shows different size or shape across imaging modalities, the probability map does not reflect the same intensity and contrast pattern shown in the component that best defines the lesion. We design the selective dense conditional

This study was supported by the Integrated Diagnostics (IDX) Program, Department of Radiological Sciences & Pathology David Geffen School of Medicine at UCLA.

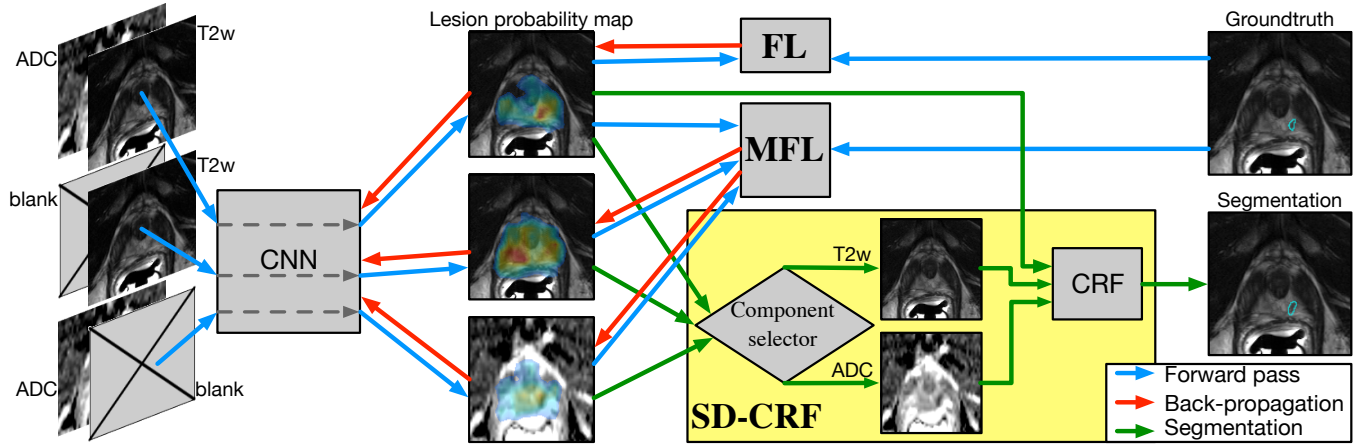


Fig. 2. The CNN for joint PCa lesion detection and segmentation. ADC and T2w images stack as different channels of the CNN input [4], and FL trains for the lesion probability map as in the top row. Also, T2w and ADC images are individually passed into the CNN, and their probability maps are trained by MFL. SD-CRF determines the specific imaging component and refines the predicted probability map for lesion segmentation using the conditional random field (CRF).

random field (SD-CRF) 1) to select a certain imaging component in which the lesion is clearly observable, and 2) to fit the predicted probability into lesion segmentation with respect to the intensity pattern of the selected imaging component, instead of simply thresholding the predicted probability map. In addition, mutual finding loss (MFL) is developed and deployed in training, to enable the imaging component selection.

Our contributions are summarized as the proposed CNN model using 1) FL to overcome imbalanced data for improved lesion detection and 2) SD-CRF to fit the CNN prediction into a specific imaging component of mp-MRI for the refined lesion segmentation. The lesion detection is achieved by finding local maxima from the pixel-level lesion probability map [2], while the segmentation is obtained by combining the probability map into the intensity pattern from mp-MRI imaging components. We train and validate the proposed CNN model using 397 pre-operative mp-MRI exams with whole-mount histopathology-confirmed lesion annotations in 5-fold cross-validation.

2. MATERIALS AND METHODS

2.1. MRI data & lesion annotation

Under IRB approval, we collected 397 pre-operative MRI exams from patients after radical prostatectomy, satisfying that 1) whole-mount histopathology is available and 2) at least one lesion with Gleason score (GS) >6 or lesion diameter ≥ 10 mm was identified in histologic examinations. All imaging was performed in 3T scanners (Siemens Healthcare) using the standardized mp-MRI protocol. T2-weighted (T2w) images and apparent diffusion coefficient (ADC) from diffusion-weighted images were used as inputs for the CNN [4, 8].

Genitourinary (GU) radiology research fellows, supervised by a senior GU radiologist, retrospectively annotated lesions in mp-MRI with whole-mount specimens available for reference as in Figure 1. Lesions with GS=6 and histologic lesion diameter <10 mm were excluded in this study since mp-MRI was reported limited detectability for those lesions [9]. In total, we have annotated 546 lesions, including 112 (20%) GS=6 lesions, 266 (49%) GS 3+4 lesions, 109 (20%) GS 4+3 lesions, and 59 (11%) GS ≥ 8 lesions.

2.2. Imaging pre-processing

For both training and validation, the intensity of T2w images was linearly normalized to $[0, 1]$ using the intensity value of bladder as the reference for the upper threshold and zero intensity as the lower threshold. Since ADC intensity value was suggestive for cancerous tissues [10], we used fixed thresholds 600 (lower) and 2200 (upper) to normalize ADC intensity. The imaging pre-processing pipeline and the CNN were built with 16-bit integers or floating-point numbers so that the numeric precision of imaging intensity was preserved. Moreover, ADC images were registered to T2w images using rigid transformation based on the coordinate information stored in imaging files. An $80\text{mm} \times 80\text{mm}$ window centered at the prostate was cropped for each case [5], and we only used slices with annotated lesions for training and validation as in [5, 11].

2.3. CNN with imbalanced data

We build the CNN using the 101-layer deep residual network [12, 13] to predict the pixel-level lesion probability map from ADC and T2w images. As in Figure 2, focal loss (FL) [7] is adapted to train for the lesion probability map from both

ADC and T2w. FL balances the loss contributed from cancerous and non-cancerous areas by adding a focal weight to the regular cross-entropy loss, such that, for each pixel,

$$\text{FL} = (1 - p)^2 y \log(p) + p^2 (1 - y) \log(1 - p), \quad (1)$$

where $p \in [0, 1]$ is the predicted lesion probability and $y \in \{0, 1\}$ is the groundtruth label. FL adjusts the penalty based on the predicted probability. E.g., for an obvious non-cancerous pixel with $p = 0.05$, the FL is only 1/400 of the corresponding cross-entropy loss; for a cancerous pixel with $p = 0.2$, the FL remain 16/25 of the cross-entropy loss. In this way, the CNN training can concentrate on cancerous or suspicious pixels.

2.4. Mutual finding loss for multi-parametric imaging

In addition to FL which trains with respect to both ADC and T2w, mutual finding loss (MFL) is designed to train for the individual imaging components of mp-MRI.

$$\text{MFL} = \frac{1}{N} \min \{ d(y \otimes f(I_{\text{ADC}}, I_{\text{T2w}}), y \otimes f(I_{\text{ADC}}, \cdot)), d(y \otimes f(I_{\text{ADC}}, I_{\text{T2w}}), y \otimes f(\cdot, I_{\text{T2w}})) \}, \quad (2)$$

where d is the L2-distance, f denotes the CNN output, \otimes is the element-wise product, and N is the number of pixels in the image. Specifically, the L2-distance between the CNN output from both imaging components and the output from either ADC or T2w alone is calculated on cancerous areas. MFL selects the individual imaging component, in which lesions are more observable, by choosing the component with the smaller L2-distance. Then, in training, MFL minimizes the L2-distance for the selected component so that the lesions can be equivalently observed from an individual component as from both components. Since MFL aims to train for the individual component, MFL does not back-propagate to the CNN output using both components as in Figure 2.

FL and MFL are combined into the total loss during the training, such that $\text{Loss} = \text{FL} + \lambda \cdot \text{MFL}$, where λ is set to the inverted fraction of cancerous pixels to balance between FL and MFL.

2.5. Selective dense conditional random field

Selective dense conditional random field (SD-CRF) is a non-parametric post-processing step to fit the CNN probability map into the intensity pattern of a specific imaging component for the lesion segmentation. SD-CRF first determines whether ADC or T2w defines the lesion better using the component selector shown in Figure 2. The component selector compares the L2-distance on cancerous areas with either of the imaging components as in (2) and chooses the component resulting in a smaller L2-distance. I.e., the selected component, $I_{\text{Sel}} = \arg \min_{c \in \{I_{\text{ADC}}, I_{\text{T2w}}\}} d(\hat{y} \otimes f_{\text{out}}, \hat{y} \otimes f_c)$,

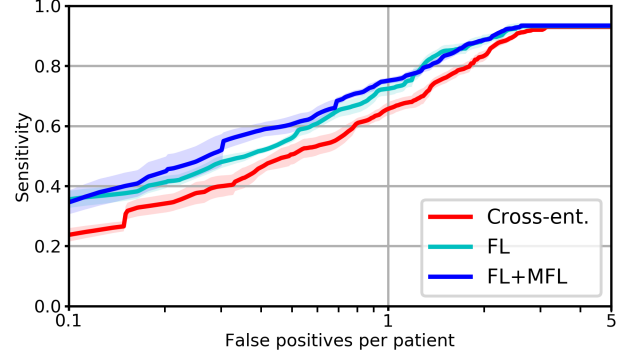


Fig. 3. The FROC analysis for lesion detection under 5-fold cross-validation. The false positives per patient (x-axis) are shown in log scale.

where f_{out} and f_c are the CNN outputs from both components and from the specific imaging component c . $\hat{y} = [f(I_{\text{ADC}}, I_{\text{T2w}}) > 0.5]$ approximates for the groundtruth y , as y is not available in testing.

Then, a conditional random field is built for the refined lesion segmentation y^* with regard to the intensity pattern in the selected imaging component and the CNN output. Specifically, as in [14], y^* is inferred by minimizing the energy E such that

$$E(y^*) = \sum_{i=1}^N \phi_u(y_i^* | I_{\text{ADC}}, I_{\text{T2w}}) + \sum_{i < j} \phi_p(y_i^*, y_j^* | I_{\text{Sel}}), \quad (3)$$

where ϕ_u is the unary potential from the negative log-likelihood of the CNN predicted probability map, and ϕ_p is the pairwise potential from i th and j th pixels. In particular, the pairwise potential is defined as $\phi_p(y_i^*, y_j^* | I_{\text{Sel}}) = -\exp(-d_{i,j}^2 - \Delta_{i,j}^2)$, where $d_{i,j}$ and $\Delta_{i,j}$ respectively are the spatial distance and the intensity difference between the i th and j th pixels.

3. RESULTS

3.1. Experiment setup

Our experiment was performed under 5-fold cross-validation. Each fold contained 317 or 318 cases for training and 79 or 80 cases for validation. We experimented four different settings in the same CNN architecture: focal loss only (FL), focal loss and mutual finding loss together (FL+MFL), FL+MFL with SD-CRF for segmentation (FL+MFL+CRF), and the regular cross-entropy loss (Cross-ent.) as the baseline. For each setting, pre-trained network weights were used as the weight initialization, and the L2-normalization with a weight of 0.0001 was added to the total loss [12]. Furthermore, common image augmentations, including random image shifting, scaling, and flipping, were applied during the training.

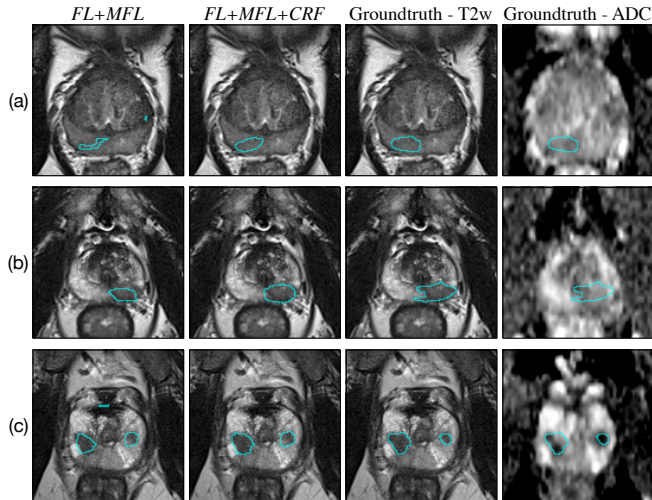


Fig. 4. Examples for lesion segmentation.

Table 1. Evaluation for lesion segmentation under 5-fold cross-validation. Numbers are reported as avg \pm std.

	Dice(%)	Dice-all(%)	HD-95 (mm)
<i>Cross-ent.</i>	36.7 \pm 1.6	18.0 \pm 1.4	7.19 \pm 0.15
<i>FL</i>	50.0 \pm 2.3	34.6 \pm 2.1	5.63 \pm 0.09
<i>FL+MFL</i>	53.8 \pm 1.8	39.7 \pm 2.0	5.24 \pm 0.15
<i>FL+MFL+CRF</i>	57.3 \pm 1.5	39.9 \pm 1.7	5.38 \pm 0.21

3.2. Lesion detection

Lesion detection is evaluated by the free-response receiver operator characteristics (FROC) analysis [2, 4]. FROC measures the lesion detection sensitivity versus the number of false positive detections per patient. Given a predicted lesion probability map, detection points are located by finding the local maxima [2]. A detection point is considered to be either true positive if it is in or within 5mm of an annotated lesion contour in the same slice [15], or otherwise false positive.

As shown in Figure 3, FL demonstrated its effectiveness in lesion detection, compared with *Cross-ent.* At 0.5 false positives per patient, *FL* and *FL+MFL* had a sensitivity of 56.0% and 60.6%, 5.5% and 10.1% higher than *Cross-ent.* Similarly, at 1 false positive per patient, *FL* and *FL+MFL* had 72.4% and 75.1% sensitivity, compared to the 65.7% sensitivity from *Cross-ent.* Moreover, for 80% sensitivity, *FL* and *FL+MFL* require 1.28 and 1.35 false positives per patient, reducing 28.7% and 25.3% of false positives from *Cross-ent.*

3.3. Lesion segmentation

Lesion segmentation is evaluated by Dice coefficient (Dice) and Hausdorff-95 distance (HD-95) defined in the brain tumor segmentation challenge [16]. Dice assesses the similarity between the predicted and groundtruth regions, and HD-95 measures the closeness of the two contour boundaries. However,

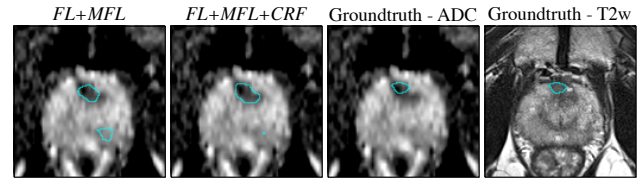


Fig. 5. Another example of lesion segmentation. The ADC image is used as the background.

compared with brain tumors, PCa lesions have highly variable detectability [17]. Thus, for the evaluation of lesion segmentation, we only focus on lesions that are roughly on target, to reduce the impact from misdetection. As in [2], we consider predicted lesions, whose centers of mass are within 10mm of centers of mass of any groundtruth lesions, as the predicted lesions on target. Similarly, the groundtruth lesions within 10mm of predicted lesions are considered as groundtruth lesions being targeted. For Dice and HD-95, we evaluate based on predicted lesions on target and groundtruth lesions being targeted, so that the various lesion detection performance (e.g., groundtruth lesions being missed or false positive predicted lesions) does not affect the evaluation for segmentation. We also include Dice-all to evaluate based on all predicted and groundtruth lesions.

The results for lesion segmentation are shown in Figure 4 and Table 1. SD-CRF improved the segmentation by refining the lesion contours and/or rejecting some noise predictions. As in Table 1, *FL+MFL+CRF* received higher Dice than *FL+MFL* and *FL*. However, *FL+MFL* had HD-95 marginally lower than *FL+MFL+CRF*. This is because all groundtruth lesion contours were annotated in T2w images, while SD-CRF sometimes fits the prediction into ADC, in which some lesions have different appearance causing this variation as in Figure 5.

4. DISCUSSION

We proposed the improved CNN model to jointly detect PCa lesions and segment lesion contours. The CNN was trained by focal loss (FL) to overcome the imbalance between normal and cancerous areas, and we designed selective dense conditional random field (SD-CRF) to refine the CNN prediction into the lesion segmentation based on the intensity pattern of a specific imaging component of mp-MRI. The improved CNN model does not limit to two imaging components and can be extended into multiple imaging components. We trained and validated the CNN under cross-validation using 397 pre-operative mp-MRI exams with groundtruth lesion contours confirmed by whole-mount histopathology. Further study would be needed to test and adapt the CNN for the heterogeneous imaging parameters as our study was based on consistent imaging parameters from a single manufacturer.

5. REFERENCES

- [1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal, "Cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [2] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman, "Computer-aided detection of prostate cancer in mri," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1083–1092, 2014.
- [3] Shijun Wang, Karen Burt, Baris Turkbey, Peter Choyke, and Ronald M Summers, "Computer aided diagnosis of prostate cancer on multiparametric mri: a technical review of current research," *BioMed research international*, vol. 2014, 2014.
- [4] Yohannes Tsehay, Nathan Lay, Xiaosong Wang, Jin Tae Kwak, Baris Turkbey, Peter Choyke, Peter Pinto, Brad Wood, and Ronald M Summers, "Biopsy-guided learning with deep convolutional neural networks for prostate cancer detection on multiparametric mri," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 642–645.
- [5] Atilla P Kiraly, Clement Abi Nader, Ahmet Tuysuzoglu, Robert Grimm, Berthold Kiefer, Noha El-Zehiry, and Ali Kamen, "Deep convolutional encoder-decoders for prostate cancer detection and classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 489–497.
- [6] Baris Turkbey, Haresh Mani, Omer Aras, Ardeshtir R Rastinehad, Vijay Shah, Marcelino Bernardo, Thomas Pohida, Dagane Daar, Compton Benjamin, Yolanda L McKinney, et al., "Correlation of magnetic resonance imaging tumor volume with histopathology," *The Journal of urology*, vol. 188, no. 4, pp. 1157–1163, 2012.
- [7] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.
- [8] Jeffrey C Weinreb, Jelle O Barentsz, Peter L Choyke, Francois Cornud, Masoom A Haider, Katarzyna J Macura, Daniel Margolis, Mitchell D Schnall, Faina Shtern, Clare M Tempny, et al., "Pi-rads prostate imaging-reporting and data system: 2015, version 2," *Eur. Urol.*, vol. 69, no. 1, pp. 16–40, 2016.
- [9] Nelly Tan, Daniel J Margolis, David Y Lu, Kevin G King, Jiaoti Huang, Robert E Reiter, and Steven S Raman, "Characteristics of detected and missed prostate cancer foci on 3-t multiparametric mri using an endorectal coil correlated with whole-mount thin-section histopathology," *American Journal of Roentgenology*, vol. 205, no. 1, pp. W87–W92, 2015.
- [10] Chiho Sato, Shinji Naganawa, Tatsuya Nakamura, Hisashi Kumada, Shunichi Miura, Osamu Takizawa, and Takeo Ishigaki, "Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate," *Journal of Magnetic Resonance Imaging*, vol. 21, no. 3, pp. 258–262, 2005.
- [11] Zhiwei Wang, Chaoyue Liu, Danpeng Cheng, Liang Wang, Xin Yang, and Kwang-Ting Cheng, "Automated detection of clinically significant prostate cancer in mp-mri images based on an end-to-end deep neural network," *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1127–1139, 2018.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [13] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [14] Philipp Krähenbühl and Vladlen Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.
- [15] Alan Priestler, Shyam Natarajan, Pooria Khoshnoodi, Daniel J Margolis, Steven S Raman, Robert E Reiter, Jiaoti Huang, Warren Grundfest, and Leonard S Marks, "Magnetic resonance imaging underestimation of prostate cancer geometry: use of patient specific molds to correlate images with whole mount pathology," *The Journal of urology*, vol. 197, no. 2, pp. 320–326, 2017.
- [16] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993, 2015.
- [17] Oliver Ruprecht, Philipp Weisser, Boris Bodelle, Hanns Ackermann, and Thomas J Vogl, "Mri of the prostate: interobserver agreement compared with histopathologic outcome after radical prostatectomy," *European journal of radiology*, vol. 81, no. 3, pp. 456–460, 2012.