Joint Detection and Diagnosis of Prostate Cancer in Multi-parametric MRI Based on Multimodal Convolutional Neural Networks

Xin Yang[1(✉)], Zhiwei Wang[1], Chaoyue Liu[1], Hung Minh Le[1], Jingyu Chen[1], Kwang-Ting (Tim) Cheng[2], and Liang Wang[3(✉)]

[1] School of EIC, HUST, Wuhan, China
xinyang2014@hust.edu.cn
[2] School of Engineering, HKUST, Kowloon, Hong Kong
[3] Department of Radiology, Tongji Hospital, HUST, Wuhan, China
1311935212@qq.com

Abstract. This paper presents an automated method for jointly localizing prostate cancer (PCa) in multi-parametric MRI (mp-MRI) images and assessing the aggressiveness of detected lesions. Our method employs multimodal multi-label convolutional neural networks (CNNs), which are trained in a weakly-supervised manner by providing a set of prostate images with image-level labels without priors of lesions' locations. By distinguishing images with different labels, discriminative visual patterns related to indolent PCa and clinically significant (CS) PCa are automatically learned from clutters of prostate tissues. Cancer response maps (CRMs) with each pixel indicating the likelihood of being part of indolent/CS are explicitly generated at the last convolutional layer. We define new back-propagate error of CNN to enforce both optimized classification results and consistent CRMs for different modalities. Our method enables the feature learning processes of different modalities to mutually influence each other and, in turn yield more representative features. Comprehensive evaluation based on 402 lesions demonstrates superior performance of our method to the state-of-the-art method [13].

Keywords: Prostate cancer detection and diagnosis · Convolutional neural network · Multimodal fusion

1 Introduction

Early detection, diagnosis and treatment of prostate cancer (PCa) are critical for increasing the survival rate of patients. The mp-MRI has been recently demonstrated to be effective for PCa detection and risk assessment [2,7]. However, interpreting mp-MRI sequences manually requires substantial expertise and labor from radiologists, and usually results in low sensitivity and specificity. Therefore, automated PCa detection and diagnosis from mp-MRI would be of high value.

M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 426–434, 2017.
DOI: 10.1007/978-3-319-66179-7_49

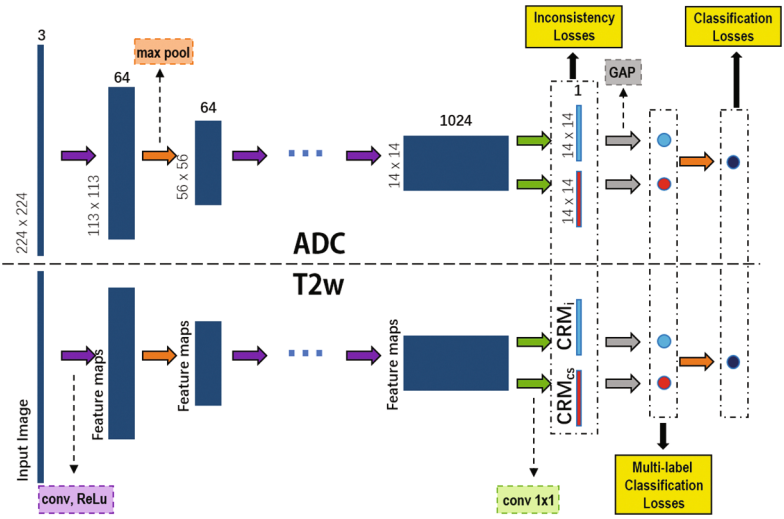

Fig. 1. Architecture of our multimodal CNNs for joint PCa detection and diagnosis.

Several studies [1–4,10,11] have been made in the past decade. In general, existing methods for PCa detection and diagnosis rely on set of separate steps which typically include voxel-level classification to generate candidate lesions from entire image sequences followed by region-level classification for verification and Gleason score (GS) grading. Existing methods differ with each other in terms of features used for representing voxels and candidate regions, MRI modalities, and methods used for classification. For instance, for feature representation Litjens et al. [4] represented each voxel using intensities and blobness of apparent diffusion coefficients (ADC), homogeneity, texture strength, etc. In [7], Peng et al. represented manually selected candidate regions using the 10th percentile and average ADC values, intensity histogram skewness of T2 weighting imaging data (T2w), etc. In [2], Fehr et al. represented regions using the first- and second-order texture features computed from the T2w and ADC images. For classification, the authors in [1] proposed cost-sensitive SVMs integrated with conditional random fields for voxel classification. In [5], Niaf et al. introduced a probabilistic SVM to address the problem in region classification when the target data include some uncertainty information. For multimodal fusion, Tiwari et al. [10] designed a semi-supervised multi-kernel graph embedding method for fusing structural and metabolic imaging data from mp-MRI. However, the reliance on separate, and sequentially executed, classification steps of these existing methods could lead to an unsatisfactory sensitivity as mis-detection of cancerous tissues in early steps can be hardly recovered in later steps. In addition, features used in each classification step are empirically designed, and hence their performance for a large-scale application with high data variation remains unclear. Moreover, existing methods either directly concatenate handcrafted features extracted from different modalities or combined multimodal results using weighted summation [12],

the relevance among multimodal features and the methods for effectively fusing multimodal information from mp-MRI have not been investigated.

This paper presents an automated method which can jointly localize PCa in mp-MRI images and assess the aggressiveness of detected lesions based on a single-stage multi-label classifier. Specifically, we train multimodal CNNs from a set of 2D prostate slices including only image-level labels indicating whether a slice containing indolent PCa, CS PCa or no PCa. The architecture of our CNNs (as shown in Fig. 1) is based on GoogLeNet [9] with three important modifications: (1) We replace fully connected layers (FC) layers of GoogLeNet with a convolutional (Conv) layer which enables us to explicitly generate two cancer response maps (CRMs) for indolent and CS cancers respectively for each modality. Each pixel on a CRM indicates the likelihood of this location be part of indolent/CS; (2) We apply the global average pooling (GAP) operation to CRMs to explicitly search for locations with large responses which indicate representative PCa patterns contribute largely for classifying the image as indolent and/or CS; (3) We design back-propagate error of CNN which not only can model multiple classes (i.e. indolent and CS) present in an image, but also can better fuse multimodal information of an mp-MRI sequence by enforcing consistent CRMs from both modalities, i.e. ADC and T2w. Such enforcement enables the feature learning process of both modalities mutually affected by each other for producing consistent maps, yielding consistent and discriminant visual features relevant to indolent and CS PCa. Extensive experiments on 402 pathologically-validated lesions demonstrate that our method achieves detection rate and precision of 80.0% and 83.7% for indolent PCa, and 84.4% and 90.5% for CS lesions, which is superior to the state-of-the-art method [13].

2 Our Method

We utilized two commonly used modalities of mp-MRI, i.e. T2w and ADC since they can provide complementary information and their fusion can effectively improve accuracy of PCa detection and diagnosis. Before inputting the images to the CNN, we register the T2w and ADC sequences using non-rigid registration based on mutual information and normalize the intensity of each T2w slice and ADC slice to 0–255 to reduce intensity variations among patients. We input each registered and normalized ADC-T2w pair into our multimodal CNNs and obtain the PCa locations and the aggressiveness of each lesion. We follow a typical clinical practice to grade the PCa aggressiveness into three classes: noncancerous, indolent PCa (GS $\leq$ 6), and CS PCa (GS $\geq$ 7). In the following, we present details of our multimodal CNNs.

2.1 Weakly-Supervised Multimodal CNNs

We first describe how to localize PCa and assess aggressiveness of each lesion for a single modality followed by multimodal fusion and details of training.

PCa Localization and Diagnosis based on Single MRI Modality. Our CNN architecture is modified from GoogLeNet [9] by first replacing the FC

layers of GoogLeNet with a Conv layer. Specifically, at newly added Conv layer (denoted by green arrows in Fig. 1), all feature maps are convolved with two convolutional kernels of $(1024 \times 1 \times 1)$, yielding two feature maps of $(1 \times 14 \times 14)$, i.e. CRM_i and CRM_{cs}, for a single modality. We apply the GAP operation to each feature map to convert it into a single image-level GAP score. It is desirable that the output of CNN produces three image-level scores p_N, p_i and p_{cs} representing the probability of this slice not containing PCa, containing indolent PCa and CS PCa, respectively. As a slice could contain multiple lesions including both indolent and CS lesions, the task of identifying the presence of indolent and/or CS PCa is a multi-label classification problem, which can be implemented using separate binary classifier. Accordingly, we define the multi-label classification loss l_{multi} by summing up the two binary logistic regression losses,

$$l_{multi} = log(1 + e^{-y_i f_i}) + log(1 + e^{-y_{cs} f_{cs}}) \tag{1}$$

where f_i and f_{cs} are GAP scores for indolent and CS respectively and y_i and $y_{cs} \in \{-1, 1\}$ are image-level labels indicating the absence/presence of indolent and CS lesions in a slice. We further project f_i and f_{cs} to a value between 0 and 1 by the sigmoid function so as to generate the probability p_i and p_{cs}. Since a slice could either contain PCa or not contain PCa, classification of noncancerous vs. cancerous can be treated as a logistic regression problem. The classification loss function l_{cancer} is therefore a cross entropy error function,

$$l_{cancer} = -\left[p\ log(y) + (1 - p)\ log(1 - y)\right] \tag{2}$$

where $y \in \{0, 1\}$ is the image-level label indicating the absence/presence of PCa in a slice, and p is the probability of this slice to be classified as cancerous which is calculated by first applying the max pooling operation to f_i and f_{cs} and then project the $max\{f_i, f_{cs}\}$ to $[0, 1]$ by the sigmoid function. Accordingly, p_N is equal to $(1 - p)$.

Since the prediction p_i is obtained by directly averaging all entries of CRM_i, when conducting the back propagation (BP) algorithm, CNN weights will be updated to suppress units in CRM_i for normal slices and slices containing only CS lesions and meanwhile activate units in CRM_i whose receptive fields are discriminative visual patterns for slices containing indolent PCa. Similar explanation is also applicable to CRM_{cs}, in which regions containing CS PCa relevant patterns will be emphasized during the BP procedure.

Multimodal Fusion. As lesions should appear at the same locations in both T2w and spatially aligned ADC slices, the CRMs of an ADC slice should be consistent with those of a T2w slice. However, training two CNN models independently for ADC and T2w images (i.e. CNN_{ADC} and CNN_{T2w}) cannot guarantee consistency between the two modalities. As shown in Figs. 2(c) and (d), the CRM_{cs} obtained based on CNN_{T2w} is inconsistent with that of ADC. Additionally, pixels related to CS PCa are not correctly highlighted on neither T2w nor ADC slices. We believe the inconsistency and the false responses are due to weakly-supervised learning, which guides CNN to 'see' not only PCa-relevant patterns but also irrelevant visual patterns. To address this problem, we enforce

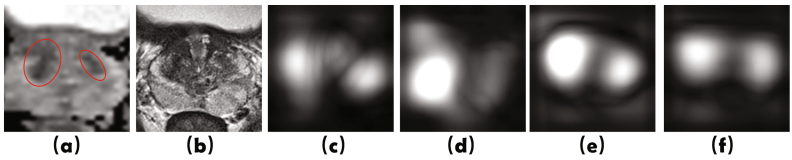

Fig. 2. (a) and (b) Registered ADC and T2w images from the original data, the red circles indicate two CS lesions. (c) and (d) are CRMs based on CNN_{ADC} and CNN_{T2w}. (e) and (f) are CRMs for ADC and T2w based on our multimodal CNNs.

the CNN models of ADC and T2w to generate consistent CRMs by defining a normalized inconsistency loss function,

$$l(CRM_{ADC}, CRM_{T2w}) = \frac{1}{2N}(\|\sigma(CRM_{i\_ADC}) - \sigma(CRM_{i\_T2w})\|^2 + \|\sigma(CRM_{cs\_ADC}) - \sigma(CRM_{cs\_T2w})\|^2) \quad (3)$$

where $\sigma(\cdot)$ is the sigmoid function and $\|\sigma(CRM_{.\_ADC}) - \sigma(CRM_{.\_T2w})\|$ calculates the differences between the maps for indolent/CS PCa of ADC and those of T2w, N is total number of pixels. The value of $l(CRM_{ADC}, CRM_{T2w})$ is a value between 0 and 1. Given the five loss functions, the back-propagated error E is defined as the weighted sum of the losses:

$$E = w_1(l_{multi\_ADC} + l_{multi\_T2w}) + w_2(l_{cancer\_ADC} + l_{cancer\_T2w}) + w_3(l(CRM_{ADC}, CRM_{T2w})) \quad (4)$$

where w_1, w_2 and w_3 are weights for adjusting the contributions of the five items to E for an optimized CNN model. In our experiments we set $w_1 = w_2 = 1, w_3 = 0.2$. Figures 2(e) and (f) show that the CRMs obtained based on our fusion model are much more consistent. In addition, the true responses at CS lesion regions are better highlighted and false responses due to irrelevant visual patterns are suppressed.

Training. We utilized a pre-trained CNN model from [13]. We fine-tuned the Conv layers of GoogLeNet, and trained the newly added Conv layer from scratch using 10,791 pairs of registered ADC-T2w images. Specifically, among all the training images, 1,379, 499, 474 and 17 image pairs are from the original mp-MRI data of patients without prostate cancers, with indolent lesions only, with CS lesions only and with both types of lesions, respectively. We augmented original data based on non-rigid image deformation as suggested by [8] to simulate another 8,422 samples, increasing both the data amount and data variety for CNN training.

2.2 Post-processing

Based on statistical analysis, we observed that CRM_{ADC} consistently provides better localization accuracy than CRM_{T2w} for both indolent and CS lesions,

thus we use CRM_{ADC} for localization. Specifically, we first adopt 'shift&stitch' to upscale the CRM_{ADC} to the size of 224×224 to minimize information loss. Second, we perform non-maximum suppression on CRM_{ADC} to detect local maximum points as the candidates followed by adaptive thresholding based on Otsu [6] to excluded false positives.

3 Experiment Results

3.1 Experimental Setup

Patient Characteristics: The study was approved by the local institutional review board. We collected mp-MRI data from 364 patients pathologically-validated by a 12-core systematic TRUS-guided plus targeted prostate biopsy. Among all validated lesions, there are 264 indolent lesions and 138 CS lesions. We sampled images including normal/benign tissues from 88 were BPH patients.

Reference Standard: For each patient who was diagnosed as PCa based on TRUS-guided prostate biopsy, two experts manually outlined the corresponding regions of interest (ROIs) on T2w slices based on the biopsy-proven locations. The reference standard of lesion regions was intersections of the two manual delineations of the two experts.

Evaluation Metrics: We evaluate the performance of image-level classification using four metrics, namely, area under curve (AUC), sensitivity (Sensi), specificity (Speci) and accuracy (Accur). We evaluate the performance of indolent/CS PCa localization in lesion-contained images using recall, precision, F1-score and accuracy. Specifically, we count the numbers of successfully localized indolent lesion (LL_i) and CS lesions (LL_{cs}), the number of localized CS lesions while incorrectly classified as indolent (FL_i) and the number of localized indolent lesions while incorrectly classified as CS (FL_{cs}). Accordingly, the localization accuracy of indolent and CS PCa is respectively calculated as $LL_i/(N_i + FL_{cs})$ and $LL_{cs}/(N_{cs} + FL_i)$, where N_i and N_{cs} are the total number of indolent PCa lesion and CS PCa lesion respectively.

3.2 Results

Image-level Classification. We first evaluate the performance of our method for distinguishing: (i) slices containing PCa from those not containing PCa (*Cancer vs. Noncancer*), (ii) slices containing only indolent PCa from those containing no indolent PCa (*Indolent vs. Nonindolent*), and (iii) slices containing only CS PCa from those containing no CS (*CS vs. NonCS*). We compare our method with two single-modality versions of our method based on ADC only (CNN_{ADC}) and T2w only (CNN_{T2w}), respectively. In addition, we also compare our method with a state-of-the-art CNN network [13] for class-specific object localization based on image-level training. Note that the network in [13] was originally trained to classify 1000 classes. For a fair comparison, we replaced their 1000-node last layer

with the last two classification layers of our network, fine-tuned their networks using the same set of ADC and T2w training images, and then combined ADC and T2w by concatenating features from each network of a single modality. The modified version of [13] is denoted as CNN_{concat}. Table 1 shows comparison of results among four methods. Two conclusions can be drawn from the results: (1) combining information of both modalities could improve the overall performance (i.e. AUC and accuracy) for all three tasks, (2) our fusion method achieves superior performances to the method directly concatenating features.

Indolent/CS PCa localization in Images Containing Lesions. We evaluate the performance of indolent/CS PCa localization in lesion-contained images based on CRM_i and CRM_{cs} respectively. We also compare our method with CNN_{ADC}, CNN_{T2w} and CNN_{concat}. Table 2 shows that the overall localization performance of our method, indicated by F1-score and accuracy, is significantly better than the single-modality CNN models and CNN_{concat}. For some cases better precision values are achieved by CNN_{concat} than ours while its the recall is extremely poor, which results in an unsatisfactory overall performance. Comparing the sensitivity values in Table 1 with the corresponding recall values in Table 2, we observe that the other three methods may mis-detect many true indolent and CS lesions even though they can correctly identify the presence of indolent/CS PCa. We believe the reason for these results is because without proper mutual guidance from both modalities in the CNN feature learning process, CNN_{ADC}, CNN_{T2w} and CNN_{concat} 'see' many PCa-irrelevant visual patterns and incorrectly rely on those patterns for image-level classification. In contrast, our well-designed loss functions help CNNs overcome the problem mentioned above and achieve consistent performance for both classification and localization.

Table 1. Performance of image-level classification

	Cancer vs. Noncancer				Indolent vs. Nonindolent				CS vs. NonCS			
	AUC	Accur	Sensi	Speci	AUC	Accur	Sensi	Speci	AUC	Accur	Sensi	Speci
CNN_{ADC}	0.995	0.970	0.960	0.982	**0.957**	0.888	0.866	0.900	**0.978**	**0.933**	0.888	**0.958**
CNN_{T2w}	0.982	0.940	0.920	0.966	0.905	0.837	0.844	0.833	0.972	0.918	0.800	0.977
CNN_{concat}	0.911	0.970	**0.986**	0.950	0.946	0.881	**0.911**	0.866	0.968	0.925	0.933	0.922
Ours	**0.998**	**0.985**	**0.986**	**0.983**	**0.957**	**0.896**	0.844	**0.922**	**0.978**	**0.933**	**0.956**	0.922

Table 2. Performance of Indolent/CS PCa localization in lesion-contained images

	Indolent PCa				CS PCa			
	Recall	Precision	F1-score	Accuracy	Recall	Precision	F1-score	Accuracy
CNN_{ADC}	0.644	0.903	0.75	0.569	0.711	0.842	0.75	0.667
CNN_{T2w}	0.467	0.840	0.60	0.438	0.644	0.906	0.75	0.592
CNN_{concat}	0.489	**0.957**	0.65	0.455	0.689	**0.968**	0.81	0.681
Ours	**0.800**	0.837	**0.82**	**0.729**	**0.844**	0.905	**0.87**	**0.720**

4 Conclusion

This paper presents an automated method for jointly localizing PCa in mp-MRI images and assessing the aggressiveness of the detected lesions. Our method employs multi-label CNNs to automatically learn representative features relevant to indolent/CS PCa. To further enhance the performance of our CNNs, information of ADC and T2w are combined by enforcing consistent cancer response maps from different modalities in the CNN feature learning process, which can help to generate more consistent and discriminative PCa features. Extensive experiments on a large dataset demonstrate superior performance of our method to the state-of-the-art method [13].

Acknowledgment. This work is funded by National Natural Science Foundation of China: 61502188.

References

1. Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Trachtenberg, J., Yetik, I.S.: Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. TIP **19**(9), 2444–2455 (2010)
2. Fehr, D., Veeraraghavan, H., Wibmer, A., Gondo, T., Matsumoto, K., Vargas, H.A., Sala, E., Hricak, H., Deasy, J.O.: Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images. Proc. NAS **112**(46), E6265–E6273 (2015)
3. Lemaitre, G.: Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging. Ph.D. thesis, Universite de Bourgogne; Universitat de Girona (2016)
4. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: Computer-aided detection of prostate cancer in MRI. TMI **33**(5), 1083–1092 (2014)
5. Niaf, E., Flamary, R., Rouviere, O., Lartizien, C., Canu, S.: Kernel-based learning from both qualitative and quantitative labels: application to prostate cancer diagnosis based on multiparametric MR imaging. TIP **23**(3), 979–991 (2014)
6. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**(285–296), 23–27 (1975)
7. Peng, Y., Jiang, Y., Yang, C., Brown, J.B., Antic, T., Sethi, I., Schmid-Tannwald, C., Giger, M.L., Eggener, S.E., Oto, A.: Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with gleason scorea computer-aided diagnosis development study. Radiology **267**(3), 787–796 (2013)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
9. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceeding of CVPR, pp. 1–9 (2015)

10. Tiwari, P., Kurhanewicz, J., Madabhushi, A.: Multi-kernel graph embedding for detection, gleason grading of prostate cancer via MRI/MRS. Med. Image Anal. **17**(2), 219–235 (2013)
11. Wang, S., Burtt, K., Turkbey, B., Choyke, P., Summers, R.M.: Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. In: BioMed Research International 2014 (2014)
12. Xinran, Z., HungLe, M., Holden, W., Michael, K., Steven, R., William, H., Xin, Y., Kyunghyun, S.: Fine-tuned deep convolutional neural network for automatic detection of clinically significant prostate cancer with multi-parametric MRI. In: Proceeding of ISMRM (2017, to appear)
13. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceeding of CVPR (2016)