



A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images

Ruba Alkadi¹ · Fatma Taher¹ · Ayman El-baz² · Naoufel Werghi¹

© Society for Imaging Informatics in Medicine 2018

Abstract

We address the problem of prostate lesion detection, localization, and segmentation in T2W magnetic resonance (MR) images. We train a deep convolutional encoder-decoder architecture to simultaneously segment the prostate, its anatomical structure, and the malignant lesions. To incorporate the 3D contextual spatial information provided by the MRI series, we propose a novel 3D sliding window approach, which preserves the 2D domain complexity while exploiting 3D information. Experiments on data from 19 patients provided for the public by the Initiative for Collaborative Computer Vision Benchmarking (I2CVB) show that our approach outperforms traditional pattern recognition and machine learning approaches by a significant margin. Particularly, for the task of cancer detection and localization, the system achieves an average AUC of 0.995, an accuracy of 0.894, and a recall of 0.928. The proposed mono-modal deep learning-based system performs comparably to other multi-modal MR-based systems. It could improve the performance of a radiologist in prostate cancer diagnosis and treatment planning.

Keywords Magnetic resonance imaging · Prostate cancer · Deep convolutional encoder-decoder

Introduction

On a worldwide scale, prostate cancer (CaP) has been reported as the second most diagnosed cancer and the fifth leading cause of cancer deaths in men. In 2012, an estimated 1.1 million men were diagnosed with prostate cancer worldwide [8]. Several screening and diagnostic tests are carried out in daily clinical routines to ensure early detection and treatment. Particularly, an increased Prostate Specific

Antigen (PSA) level is usually followed by MR screening, and an ultra-sound (TRUS)-guided biopsy, respectively. Besides its noninvasive nature, magnetic resonance (MR) screening has shown high potential in early diagnosis, monitoring, and treatment planning of prostate cancer. Interestingly, the detection rate of prostate cancer in MRI has been recently reported to be in the range of 0.44 to 0.87, which is higher than the reported finding of blind TRUS biopsy making it a suitable noninvasive alternative [9]. Other key advantages of MR screening lies in its ability to provide information about the tumors' location, volume, and level of malignancy. These characteristics are especially essential for the active surveillance, where indolent lesions that are confined in the prostate are regularly monitored to ensure that cancer cells are not developing in an irregular manner that necessitates medical intervention or other therapies [1].

Basically, there are four main MRI modalities that are used in CaP diagnosis. These include T_2 weighted (T2W), dynamic contrast enhanced (DCE), diffusion weighted (DW), and magnetic resonance spectroscopy (MRS). T2W MRI is the basic MRI modality that uses the transverse relaxation time T_2 to construct a grayscale image of the scanned object. Due to its increasing popularity and availability by many health providers, T2W-MR images have been an effective tool for noninvasive CaP diagnosis

<http://i2cvb.github.io/>

✉ Ruba Alkadi
ruba.alkadi@ku.ac.ae

Fatma Taher
fatma.taher@ku.ac.ae

Ayman El-baz
ayman.elbaz@louisville.edu

Naoufel Werghi
naoufel.werghi@ku.ac.ae

¹ Khalifa University of Science and Technology, PO Box 127788, Abu Dhabi, United Arab Emirates

² University of Louisville, Louisville, KY 40292, USA

[30]. The main advantage of this modality is that it allows to visually differentiate between normal prostatic tissues and cancerous tissues by means of intensity and homogeneity [25, 44]. More precisely, malignant tissues are characterized by lower signal intensity in the peripheral zone (PZ) of the prostate and a more homogeneous appearance in both the central gland (CG) and the PZ compared to the surrounding healthy tissues [44]. This mainly results from the presence of a foci of congested glands (cancer) surrounded by less dense benign cells, which in turn translates to a region of low signal intensity in the image. The other main advantage of this modality is that it encodes details of the zonal anatomy of the prostate gland. That is, the CG is well-distinguished from the PZ of the prostate and the surrounding non-prostatic tissues [19, 25].

In practice, MRI images are interpreted by an experienced radiologist who produces a report of findings for each case. Typically, a radiologist spends 14–15 years of post-high-school education and training before commencing his work as a radiologist [27]. Because of the well-structured nature of radiologists' work which mainly relies on image analysis and interpretation, limited interaction exists between the radiologist and patients. Particularly, this sufficiently isolated and independent nature of work makes it an attractive candidate for computerized image processing. Mazurowski et al. [27] have recently suggested that some existing computer vision algorithms could considerably: (1) reduce the efforts required by radiologists, (2) improve the quality of the interpretation by assessing new features in images that have not been previously assessed by humans, (3) improve the repeatability of the decision, and (4) reduce the time needed for image interpretation. These facts have fueled the need for developing an expert computer-aided MRI-based CaP detection system that competes well with experienced radiologists.

Despite the fact that computer aided diagnostic tools are very promising, there are several challenges associated with the employment of computer vision algorithms on medical images. In particular, MR images are grayscale 3D images that suffer from high signal to noise ratio, low soft tissue contrast, and other severe artifacts, resulting in inter and intra patient variabilities. Data imbalance is also another challenge that is associated with medical images, where the amount of samples of one class (benign) is usually much higher than the other (malignant). Finally, the lack of sufficiently large and comprehensive annotated datasets presents another obstacle towards the application of recent computer-vision algorithms that mainly depends on learning discriminative features from a large number of images. Thus, unlike common visual contexts, MR image analysis and interpretation require more sophisticated solutions that adapts well with these constraints.

In 2003, Chan et al. [3] pioneered the development of a computer-aided diagnostic tool that analyzes prostate MRI to detect and localize CaP. Afterward, several CADs were realized in the following decade [21, 22, 28, 31, 38–40, 43, 45], all of which followed a quite identical workflow. Generally, CAD systems presented in these studies start with pre-processing the input MR volumes for noise removal and intensity level standardization. Then, for multi-modal approaches, a registration step is usually carried out to align the different modalities and to compensate for patient movement during the screening process. This is followed by prostate gland segmentation to extract the volume of interest (VOI) at which following steps will take place. Common 2D and 3D features such as wavelet-based features, Haralick features, and Tamura's features, are extracted for each pixel in the prostate volume. These features are further reduced to produce a set of meaningful features relevant to the detection of malignant tissues. Eventually, pixel classification based on the extracted features is performed by the means of machine learning. Some post-processing techniques are usually adopted at the end to visualize suspected lesions. To a large extent, the selection of meaningful and effective features that can ensure high reproducibility remained limited [27].

In 2012, the breakthrough of deep learning in the area of computer vision had a radical impact on all its dependent fields as it outperformed the traditional pattern recognition approaches with a large margin in the ILSVRC challenge [35]. As a natural consequence of this advent, a gradual shift from handcrafted feature-based systems, to systems that intelligently 'learn' high-level features was realized in the medical image analysis domain [23]. Despite the impressive success of deep learning in other medical image analysis applications (e.g., breast mammography, brain MRI, and lung CT [18]), development of prostate cancer CAD systems is still lagging behind. Surprisingly, a recent comprehensive review [20] of more than 40 prostate MRI CAD systems reports no use of deep learning-based approaches for this specific application. While a more recent review [23] of deep learning in medical image analysis reports only very few attempts to employ deep architectures for the task of prostate cancer detection and diagnosis.

This work contributes in filling this gap by a thorough investigation of a unique type of convolutional neural networks (CNN), namely, the deep convolutional encoder-decoder architecture, in the context of deploying a fully automatic mono-parametric MRI CaP detection and localization system. We hypothesize that by a careful employment of deep CNN, a more accurate interpretation of MR scans becomes possible. We dedicate a full section below to highlight our original contributions and elaborate more about the related rationals.

Contribution

Our contribution is distinguished by the following aspects:

- We employ a deep encoder-decoder convolutional neural network to segment the unique texture of CaP in both PZ and CG. The other non-cancerous tissues are divided into three categories that correspond to the non-prostate tissues and the anatomical zones in the prostate, namely, the peripheral zone and the central gland. This setting enabled us to bypass the segmentation step and utilize the raw MR series as an input to the system.
- We exploit the 3D spatial information in the MR series without compromising the computational cost by sliding a 3D window that encloses three consecutive slices. Our experiments show that this approach improves the overall performance of the system, while preserving the same computational cost and complexity.
- Unlike other state-of-art CAD systems [19, 41, 46] that require multi-parametric MRI input to detect and localize CaP, we design and train our system to perform these tasks using only T2W images. Although, we are aware of the fact that other MRI modalities could provide more meaningful information and can definitely better guide the CAD system in identifying malignancies as suggested by earlier studies [30, 40, 44], we believe that the resulting overall improvement in the performance of the system is limited relative to the increased complexity associated with multi-parametric fusion. Practically, multi-parametric approaches demand non-trivial manual operations. For example, validating the registration of ADC maps with T2W scans requires tedious manual annotation of a large number of scans in both modalities. In addition, the acquisition time of T2W is significantly shorter than multi-modal acquisition (around 10 minutes versus 40 minutes). All these factors yield high costs in terms of labor and time resources in the clinical practical usage. In this work, we show that a mono-parametric CAD system performs comparably to its multi-parametric counterpart. This is specially true as the registration step, required by multi-parametric CADs, usually results in additional degradation of the overall system performance. It is also worth noting that we opt to process the conventional T2W sequence since it has the highest in-plane spatial resolution and therefore is the most crucial in tumor stage assessment compared to other modalities [9].
- Compared to other systems [17, 21, 41] that demand manual segmentation of the prostate as a preprocessing step, our system implicitly perform prostate segmentation by assigning a ‘non-prostate’ category to all background voxels.

The rest of the paper is organized as follows: “[Related Work](#)” surveys the literature for the state-of-the-art of CaP detection systems, “[Proposed Method](#)” elaborates on the proposed method, while “[Experiments](#)” describes the experiments. This is followed by the results and related discussion in “[Results](#)”. Finally, we present our concluding remarks in “[Discussion and Conclusions](#)”.

Related Work

Following the recent advances in hardware processing power, the development of medical image-based CAD systems has gained increasing popularity in the last decade. We briefly survey the literature of MRI-based computer-aided detection and localization of prostate cancer. While there have been a large volume of work on CaP detection and diagnosis systems, we focus our survey on systems that produce a cancer probability map at their outputs. To best illustrate the different methodologies, we categorize the reviewed systems into two main categories based on the approaches explained in “[Introduction](#)”. First, we review systems that followed the feature engineering approach. This is followed by a second set of systems that utilized deep-learning architectures as an alternative for feature extraction and classification.

Feature Engineering-Based CAD

Image processing techniques that relies on handcrafted features have gained popularity in the medical imaging analysis domain for a long period of time. To date, they still form the basis of CAD systems that are commercially available [23]. In fact, all of these systems follow a quite similar methodology encompassing a 3-stage process: (1) feature extraction, (2) feature reduction, and (3) classification. In this paradigm, the common trend was to extract a large number of statistical features from each voxel in the image, which are then reduced via a given feature selection methods. This is followed by a classification step in the high dimensional feature space, where a machine learning algorithm (e.g., SVM, Random Forest, etc) learns the optimal decision boundary. Examples of systems that adopted this paradigm are [22, 28, 38–40, 43, 45]. For a full overview of these systems, we refer the reader to [20].

From T2W MR sequence, Rampun et al. [14] identified a set of 215 features and grouped them in six classes. They used the correlation-based feature selection method [13] to reduce the feature space dimensionality by ranking the feature classes rather than individual features. The authors reported that Gaussian filters, Laplacian of Gaussian filters, image magnitude of the Sobel operator, and Tamura’s contrast

were among the most selected features. For classification of features, they evaluated the performance of nine popular classifiers and two meta-voting classifiers. Keeping all parameters in the default setting, the meta-vote (best 2) classifier outperformed all the other 10 alternatives with an AUC of 0.927, an accuracy of 0.855, and a sensitivity of 0.933. This classifier combined the results of the Bayesian Networks (BNets) and the alternating decision tree (ADTree) using the average probability combination rule. They suggested that these results are comparable to other multi-parametric-based CaP diagnosis systems.

Lemaitre et al. [19, 21], proposed a multi-stage multi-parametric MRI CAD system for the early detection and diagnosis of prostate cancer. In their final model, they selected 267 out of 331 features extracted from T2W, DW, DCE, and MRS sequences and used them to train a random forest classifier. They validated their system on data of 17 patients from the I2CVB dataset by generating a cancer response map (CRM) on each slice. This best setting achieved an average AUC of 0.836 on a leave-one-patient-out cross validation protocol.

On the same dataset, Trigui et al. [41] tested their proposed multi-parametric CAD system for the detection and localization of healthy, benign, and malignant tissues in the prostate. Similar to [21], they extracted features from the four MRI modalities and fused them in a random forest classifier. They extensively tested all possible combinations of features and classifiers to come up with the optimum design of the processing pipeline. The authors report that the best performing scheme achieved an error rate of 0.182 (accuracy 0.818) on a 10-fold cross validation using a random forest classifier with features from only two modalities (ADC maps and MRSI). It is worth noting that the resolution of the produced color-coded maps of their CAD system was limited by the low resolution of the MRS.

Adopting a similar pipeline, Wang and Zwiggelaar [46] developed a dictionary of 3D texon features from a 3D window around each voxel in the training set. These features included ADC values, DCE signals, and T2W intensity values. They concatenated the extracted features of each voxel in a feature vector which was eventually fed into a random forest classifier. On a leave-one-patient-out protocol, they achieved an average accuracy of 0.883 on 17 patients from the I2CVB dataset.

Deep Learning-Based CAD

Following the success of deep learning in the area of computer vision, many researches followed the new trend and proposed diverse CNN architectures that act directly on the MR raw data. In contrast to the previous category of approaches at which the extraction of discriminant features is done by human researchers, deep learning networks let

the machine *learns* the *optimal* set of features allowing for a better interpretation of patterns in medical images.

Attempting to explore the performance of a special type of CNNs on lesion segmentation, Kohl et al. [17] proposed the use of a generative adversarial network (GAN) to segment the prostate into peripheral zone (PZ), central gland (CG) and CaP. Their GAN consists of two parts: a U-Net-like segmentor which is trained to segment the three classes on a slice-wise fashion, and a discriminator which is supposed to distinguish between fake (generated) and expert (ground truth) segmentations. They report a sensitivity of 0.55 and a specificity of 0.98 based on bi-parametric MRI data from 152 patients. They used a fully annotated in-house dataset, where malignant lesions, PZ, and CG borders are delineated by a radiologist.

In the same context, Kiraly et al. [16] proposed another approach that also utilizes deep learning for slice-wise detection of prostate cancer in multi-parametric MR images. Similar to [17], they reformulated the task as a semantic segmentation problem and made use of a SegNet-like architecture [2] to detect possible lesions. They achieved an average area under the receiver operating characteristic curve (ROC) of 0.834 on data from 202 patients. This study however, was limited with respect to ground truth. That is, the available ground truth of was defined as limited set of single voxel in the MRI series. The authors expanded it to a full region by means of a 3D Gaussian kernel.

Recently, Yang et al. [50] modified the last layers and the loss function of a GoogLeNet-like architecture pretrained by [52]. Using an image level supervision setting, their dual path CNN-based system detected indolent and clinically significant cancer foci with a rough localization map in T2W images and apparent diffusion coefficient maps (derived from the DW modality). The authors presented a modified system in [49], where they added a CNN-based prostate region detection step at the beginning of the pipeline to roughly crop the region of interest before the feature extraction and classification take place. A further modification of their system was presented by Wang et al. [47]. In this final version, they added a term in the loss function to implicitly register ADC maps and T2W scans during the back-propagation process. An overlap loss term was also added to the loss function to account for inconsistency in the output CRM produced by the dual-path architecture. They validated their system on an in-house dataset of 160 patients [49]. In Wang et al. [47] however, further validation on the PROSTATEX challenge dataset was carried out. A maximum sensitivity of 0.8978 at one false positive per benign case was reported.

The work of both [16] and [17] is the most similar to our presented work as both papers utilized deep learning-based semantic segmentation for CaP detection and localization in 2D MR slices. However, our work differs in several

perspectives. First, we explicitly exploit 3D spatial contextual information to guide the segmentation process which eventually improves the overall system performance. Second, we simultaneously perform anatomical segmentation and lesion detection in MR images, whereby we associate malignancy probability estimation with each voxel. Third, unlike [16] and [17], our work relies on input data from only one modality (T₂W). We thus highlight the potential of extracting sufficiently meaningful information from a single modality. In fact, this also has a significant clinical advantage, as it reduces the time and cost associated with the screening process, and eliminates the need of contrast agent injection required during the acquisition of DCE MR images. Finally, and most importantly, this work implements a deeper convolutional architecture compared to the one used in [16] and, to the best of our knowledge, is the first to assess CNNs performance on the fully annotated I2CV benchmarking dataset

In Table 1, we report a nutshell recapitulation of the surveyed methods. [20].

Proposed Method

Our proposed approach falls under the category of semantic segmentation methods. This paradigm can be viewed as pixel-level understanding where each pixel in the image is assigned to a particular class representing a certain object. Apart from recognizing the objects in the scene, it implicitly delineates their boundaries in the image. This semantic interpretation aspect makes it different from classic segmentation which produces regions that share common attributes but without representing necessarily specific objects. While

semantic segmentation has been dealt with in the computer vision community since a while, the major success of this paradigm came with the advent of fully convolutional networks (FCN) [24] which paved the way to the subsequent deep learning approaches and variants in semantic segmentation. The taxonomy of deep learning architecture for semantic segmentation encompasses the aforementioned FCN models, context-aware models, and temporal models. These last two models incorporate architecture aspects that account for the spatial context and temporal information respectively [10].

Semantic segmentation architecture roughly encompasses an encoder and decoder networks. The encoder is a classification network that projects the input images into a features space. A standard classification architecture (e.g. vggNet) is usually employed in this task. The decoder semantically projects the features learned by the encoder back into the pixel space, through up-sampling cascading layers, thus gradually recovering the semantics details and the input image spatial dimensions. Most of the proposed semantic segmentation networks differ at the decoder architecture driven by improving the segmentation accuracy and the computational cost. The UNet network [32], which has been widely used in the medical imaging community has a ladder-like architecture. This terminology reflects the fact that the entire features maps across the encoding layers are transferred to the corresponding decoder layer where they are concatenated with its upsampled feature maps. While this mechanism allows the decoder to learn back, at each stage, relevant features that are lost when pooled in the encoder, it is more demanding in terms of memory consumption and computation complexity. In our work, constrained by our limited computational resources, rather

Table 1 Comparison of prostate segmentation results with the literature

	2D/3D	Modalities	Dataset size	Approach	Reported performance	Prostate segmentation	Registration	Validation protocol
Rampun et al. [30]	2D	Mono-modal	45	Engineered features	0.93 AUC	Manual	N/A	9-fold CV
Lemaitre et al. [19, 21]	3D	Multi-modal	17 ^a	Engineered features	0.834 AUC	Manual	Automatic (elastic)	Leave one out
Trigui et al. [41]	2D	Bi-modal	34 ^a	Engineered features	0.72 sensitivity 0.88 specificity	Manual	N/A	10-fold CV
Wang and Zwiggelaar [46]	3D	Multi-modal	17 ^a	Texton dictionary	0.884 AUC	Manual	Automatic (rigid)	Leave one out
Kohl et al. [17]	2D	Bi-modal	152	GANs	0.55 sensitivity 0.98 specificity	Implicit	N/A	4-fold CV
Kiraly et al. [16]	2D	Multi-modal	202	Convolutional encoder-decoder	0.843 AUC	Automatic (elastic)	Level-set	5-fold CV
Yang et al. [50]	2D	Bi-modal	360	Dual path CNN	0.89 sensitivity	Implicit	Implicit	5-fold CV

^aFrom the I2CVB dataset

than transferring the whole feature maps, we adopted the variant of transferring the pooling indices computed in the max-pooling step of the corresponding encoder as proposed in [2]. The pooling indices form a sparse feature maps with a quite lower size compared to the entire features maps in the UNet. This alleviates, therefore, the computation and the memory demand burden. As demonstrated in [2], this variant showed to achieve a good trade-off between segmentation accuracy and computational cost. Another aspect in our methods is that it exploits the volumetric aspect of the MRI images via a 2D multi-channel mechanism that accounts for the spatial contiguity of the slices in the MRI modality (see “MRI Data Encoding”).

Deep Convolutional Encoder-Decoder Architecture

We reformulate the problem of lesion detection and localization as a semantic segmentation task. Assuming that each lesion in the prostate encodes a unique texture, we produce a mask that assigns one of four labels to each voxel as a ground truth label as shown in Fig. 1. In this paradigm, our goal is to design a CNN architecture that will assign each voxel to one of these labels. We then utilize a deep convolutional encoder-decoder architecture to predict the labels of each pixel in the test set. The CNN architecture shown in Fig. 2 is similar to the one presented in [2]. This architecture consists of an encoder (contracting path), a corresponding decoder (expanding path) of the same size, a trainable SoftMax layer, and at the back-end a pixel classification layer. The encoder is made of thirteen convolutional layers having a structure similar to the first set of layers of the VGG16 network [36]. In each convolutional

layer, the input feature map is convolved with a set of trainable filters of size 3×3 and a stride of 1. This is followed by batch normalization, at which the activations are normalized by the following transformation:

$$\hat{x}_i = \gamma \frac{x_i - \mu_b}{\sqrt{\sigma_b^2 + \epsilon}} + \beta, \quad (1)$$

where μ_b and σ_b , are the mean and variance of each mini-batch at each channel, γ and β are the trainable scale and shift parameters, x_i , \hat{x}_i are the input and output of the batch normalization layer, and ϵ is a stability constant that avoids division by zero when the mini-batch variance approaches zero. In our experiments, we set $\epsilon = 1.0^{-5}$. Then, batch normalization is followed by a rectified linear unit (ReLU) which performs the thresholding operation ($\max(x, 0)$). These three layers are repeated two or three times before a max-pooling layer functions. Max pooling is performed on a window of size 2×2 with a stride of 2 to reduce the feature map size by a factor of 2 in each dimension.

Although the encoder shares similar structure with the VGG16 network, fully connected layers that are present in the VGG16 are removed from the encoder to impressively reduce the number of trainable parameters from 134M to 14.7M [2].

After retaining high-level feature maps at the end of the encoder, a decoder is trained to recover the original input size which is essential for voxel-wise classification. An up-sampling layer receives the indices of the maximum values in the pooling window of the corresponding max-pooling layer and produces upsampled sparse feature maps. Then, the resulting sparse feature maps are convolved by a

Fig. 1 Illustration of sliding a 3D window across the input volume. Note that the label of the middle slice is considered as the label for the 3D window. In the dataset each pixel at each slice is labeled as shown in the left

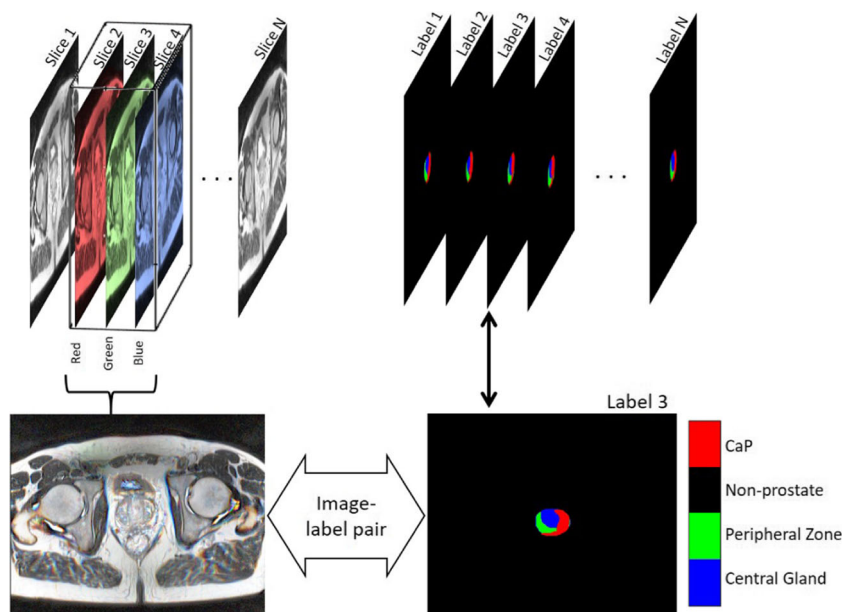
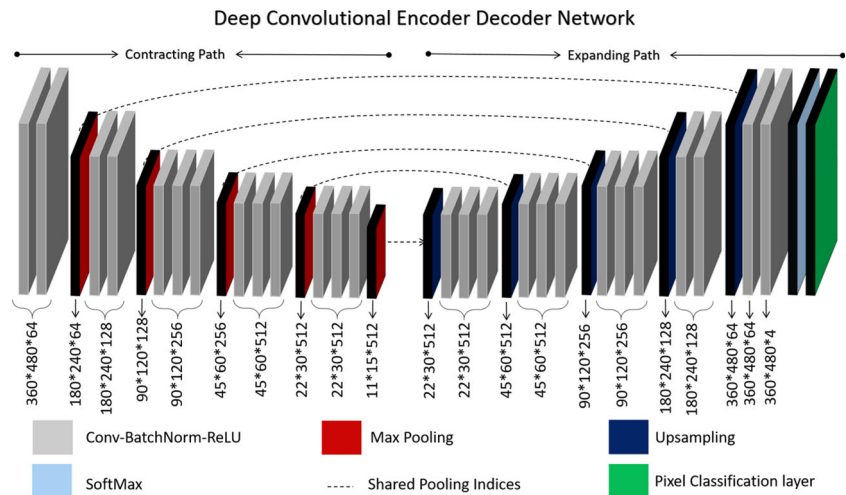


Fig. 2 The architecture of the deep convolutional encoder-decoder network presented in [2] and used to segment lesions in MRI. Note that the dimensions under each layer corresponds to the size of the output activations produced by same layer



trainable filter bank. Besides densifying the feature maps, this convolution operation reduces the depth of the input feature maps to match the depth in the corresponding encoder layers. To stabilize the inputs to the ReLU, batch normalization is also used after each convolutional layer at the decoder. This up-sampling using max-pooling indices scheme have shown to be a computationally more efficient alternative to the inverse convolution operation used in [24]. Finally, a SoftMax classifier is used to independently classify each voxel based on the resulting feature maps at the last convolutional layer. Although the architecture tends to be symmetrical, the depth of the feature maps in the outer encoder does not match its corresponding decoder depth as illustrated in Fig. 2.

For labeling purposes, a pixel classification layer is used to convert the classification probabilities that resulted from the SoftMax layer into class-labels. This layer essentially produces a label image as a final output layer.

As mentioned previously, the main advantage of this architecture as compared to other similar architectures (e.g., U-Net [32] and fully convolutional network [24]) is that pooling indices are communicated directly from the encoder layers to the corresponding decoder layers which significantly reduces the computational cost for training, without a compromise on the performance. Another advantage is that the encoder parameters of this network can be initialized by weights from VGG-16 net which is trained on the ImageNet. Finally, by utilizing the activations from the inner layers, this network enabled us not only to segment multiple lesions, but also to produce a CRM for the suspected lesions in the prostate, as will be explained later.

MRI Data Encoding

Initially, the most standard CNN architectures have been designed for 2D color images. CNN architectures for 3D

volume images in computer vision community have been approached via two paradigms namely: volumetric CNNs [48] and multi-view CNNs [29].

Volumetric CNNs had the merit to pioneer the application of 3D convolutional neural networks on voxelized shapes; however, they are criticized for the demanding cost inferred by 3D convolutions. In contrary, multi-view CNNs project the 3D images into sectional planes to obtain a sequence of 2D images which are then fed into standard 2D CNN. Generalizing well to 3D medical volume images, a more straightforward paradigm has been adopted in several works [16, 17, 47, 49, 50], whereby each 2D slice is used as a separate input of the network. However these methods do not capture 3D contextual information that are relevant to the segmentation of lesions which are volumetric in nature. Other methods such as [51] opted for volumetric CNN, and for addressing the computational cost issue, cropped the MRI into small cubes which were fed separately to the 3D network. This kind of approaches requires a stitching step to re-combine the outputs of cropped volumes.

On the other hand, Roth et al. [33, 34] proposed a multi-view CNN variant, coined 2.5D method, in which they fed three orthogonal views of the volume of interest (VOI) into the RGB channels of a deep CNN.

In our approach, in order to consider the 3D contextual information while accounting for the contiguity constraint, we propose a new representation for multi-view CNNs constructed by extending the dimension of the lowest resolution of the input MR volume into the RGB dimension. We perform this by sliding a 3D window along the MR series. Here, we use the term 2D multi-channel input to refer to a sub-volume of dimensions $x \times y \times 3$ where x and y are the slice dimensions (see Fig. 1). In both, training and testing, we feed the network by colored images that essentially encode three consecutive gray-level slices. In back-propagation, the network minimizes the error between

the label image of the middle slice and the predicted label. Hence, we expect the output of the network to be a labeled image that corresponds to the middle slice (S_m). The preceding (S_{m-1}) and following (S_{m+1}) slices are, however, used in the feature extraction forward pass to assist in extracting and learning meaningful volumetric features. To illustrate the benefit of this arrangement, consider a slice S_i that contains CaP at some region R , where $R \in S_i$. To preserve continuity of the detected lesion, there must exist at least one similar region $R' \in S_{adj}$, where $S_{adj} = S_{i-1} \cup S_{i+1}$ that satisfies $S_i(R) \cap S_{adj}(R') \neq \emptyset$. Such a correlation and other relevant 3D features are expected to be learnt by filters in the training phase, and indeed enhance the performance of the testing phase.

In principle, this approach offers three main advantages other than exploiting 3D spatial information. First, it provides a standardization technique that copes with the problem of inter-patient inconsistency in prostate gland size and thus normalizes the number of slices of the MR series. Second, it transforms gray-level slices to colored images without the need to redundantly replicating the gray image in each of the RGB channels. Finally, while being able to incorporate 3D features, this approach requires no additional computational cost compared to other volumetric CNNs.

Other challenges that are usually encountered by the application of deep learning in medical image analysis are (1) the scarcity of training samples and (2) class imbalance (i.e., classes are not fairly represented in the dataset) [11]. We addressed the first problem by standard data augmentation techniques such as addition of Gaussian noise, random reflection, and translation. On the other hand, class imbalance is usually addressed using techniques that either operate on the dataset itself or by modifying the network loss to adapt to the imbalance. The first set of techniques are more straightforward, and thus more popular. Techniques that operate on the dataset could be further categorized into two main categories, namely, minority class up-sampling and majority class under-sampling. Although several sampling techniques have been proposed in the literature [4, 14, 26, 37], very few have looked at the effects of using these techniques on deep learning models. Instead, these techniques are better investigated in machine learning approaches. The other set of techniques manipulate the learning process by explicitly quantifying the class imbalance rate. Usually, the quantified imbalance rate is added to the loss function to bias the classification towards the underrepresented class. In our proposed method, we addressed the problem of imbalanced data by reweighing each class in the loss function using the median frequency imbalance approach presented in [7]. Basically, the class

weight w_i associated to the i_{th} class c_i can be defined using the class frequency f_{c_i} as follows:

$$w_i = \text{median}(f_c)/f_{c_i}, \quad (2)$$

$$f_{c_i} = K_{c_i}/K_m, \quad (3)$$

where $f_c = \{f_{c1}, f_{c2}, \dots, f_{ck}\}$ is the set of class frequencies of all categories present in the dataset, k is the number of classes and K_{c_i} and K_m are defined as:

$$K_{c_i} = \sum_{j=1}^k n_j(c_i), \quad (4)$$

$$K_m = \sum_{j=1}^k x_j y_j, \quad (5)$$

where $n_j(c_i)$ is the total number of voxels from class c_i in slice j , and x_j and y_j are the dimensions of each training slice j . This implies that underrepresented classes are assigned a weight $w_c > 1$. Simply, the cross-entropy loss is re-weighted by the new weights of each class which is described as:

$$CE(y_i, z_i) = - \sum_{j=1}^k w_j \{y_i = j\} \log \frac{e^{z_j}}{\sum_{l=1}^k e^{z_l}} \quad (6)$$

where y_i is the ground truth label and z_i is the predicted label of the i_{th} voxel.

Experiments

Patient Characteristics and Reference Standard

We performed our experiments on a subset of a public dataset released in [20] as part of (I2CVB) Benchmark dataset. The MR scans are acquired from a cohort of patients with higher-than-normal level of PSA. All patients were screened using a 3 Tesla whole body MRI scanner (Siemens Magnetom Trio TIM, Erlangen, Germany). Four different modalities were available for each patient in the dataset; however, only the T2W sequence was used in this work, for the reasons explained earlier in “Introduction”. The dataset is composed of a total of 19 patients of which 17 have biopsy proven prostate cancer and 2 are healthy with negative biopsies. From those 17, 12 cases have malignant lesions in the PZ, 3 have malignant lesions in the CG, and 2 have invasive CaP in both PZ and CG. The mean age of the patients in this subset was 63.2 ± 9.3 years, ranging from 40 years to 82 years old.

An experienced radiologist segmented the prostate organ on T2W MR images, as well as the prostate zones (i.e., PZ

and CG), and CaP. Unlike the ProstateX dataset released by [22], the ground truth in this dataset provides full segmentation of prostate gland, its anatomical regions and the spatial extend of the cancerous lesions, making it more suitable for our application.

MRI Acquisition Protocol and Pre-processing

Three-dimensional T2W fast spin-echo (TR 3600 ms, TE 143 ms, ETL 109, slice thickness 1.25 mm) images are acquired in an oblique axial plane. The nominal matrix and field of view (FOV) of the 3D T2W fast spin-echo images are 320mm×256mm and 280 mm×240 mm, respectively, thereby offering sub-millimetric pixel resolution within the imaging plane. Consequently, the resulted volume resolution varied between $384 \times 308 \times 64$ and $448 \times 368 \times 64$ voxels.

One of the common challenges in MRI processing is the inter-patient intensity variability. Thus, a preprocessing step was inevitable to remove possible artifacts which are mainly introduced by magnetic field inhomogeneity and endo-rectal coil placement. The N4Bias correction technique [42] have been widely used in handcrafted features methods to ensure stable and robust feature extraction from the MRI modality. In our case, we are employing an end-to-end deep learning architecture which has the intrinsic capacity to accommodate, through learning, to the variability factors in the input images. Therefore, we rather used the basic histogram normalization method to improve the image contrast.

Evaluation Metrics

A fair comparison with the state-of-art CAD systems is usually challenging due to the use of different datasets and evaluation measures. In this work, we try our best to overcome this limitation by testing our system on a public dataset and assessing its performance using a variety of evaluation measures adopted in the literature.

In this setting, we use the well known measure which is the accuracy. It is defined as the ratio of the number of correctly classified voxels over the total number of voxels in a testing volume. In our case, we average the accuracy over all the testing images. In fact, the accuracy measure sometimes overestimates the system performance especially when the number of negative samples (i.e., non-cancerous voxels) is much larger than that of the positive samples (i.e., cancer voxels).

Second, the intersection over union (IoU) is used to evaluate the segmentation accuracy of the system. It is calculated by dividing the number of true positives (i.e., correctly identified cancer voxels) over the total number of true positives, false positives, and false negatives. For a specific system, this metric is usually lower than the accuracy

since it does not take into account the correctly identified negative samples.

Third, the recall is used as a measure of the ability of the CAD system to identify cancerous regions. It is defined as the number of correctly identified cancerous voxels over the number of all voxels labeled as cancer in the ground truth.

Fourth, the dice similarity coefficient (DSC) is a common measure that is used to evaluate the segmentation performance of the system. We use it here to evaluate and compare the performance of the system in segmenting the prostate gland from the surrounding tissues. As an extension to the DSC, we use the boundary F1 score (BF score) which is a modified version of the DSC score. Particularly, this measure tends to correlate better with human qualitative assessment as suggested by [5].

Finally, to be able to compare with Lemaitre et al. [21] and Wang and Zwiggelaar [47], we performed a leave-one-patient-out cross validation on the I2CVB dataset. We then computed the area under the receiver operating characteristic curve (AUC). This value is derived from the receiver operating characteristics (ROC) curve which is obtained by varying the discriminative threshold of the pixel classifier (i.e. pixel classification layer layer in our case). To obtain the ROC curve (and eventually the AUC), we utilized the activations of the SoftMax layer to produce a probability maps of the cancerous lesions. We then used these probabilities to plot the ROC curve for each lesion contained slice and finally, averaged all the curves over the test set.

A summary of these four evaluation metrics is provided in Table 2.

Training Parameters

We trained our system to be able to segment the prostate from the surrounding tissues, segment the two anatomical zones in the prostate, and produce a CRM to localize possible lesions. The network was trained using an NVIDIA Quadro GPU. A total of 2356 multi-channel slices were extracted from the dataset. The samples were split as follows: 60% (1413 slices) were used for training, 10% (236 slices) were used for validation, and the remaining 30%

Table 2 Definition of evaluation metrics (N_{TP} = number of true positives, N_{TN} = number of true negatives, N_{FP} = number of false positives, N_{FN} = number of false negatives)

Metric	Formula
Accuracy	$\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}}$
IoU	$\frac{N_{TP}}{N_{TP}+N_{FN}+N_{FP}}$
Recall	$\frac{N_{TP}}{N_{TP}+N_{FN}}$
DSC	$\frac{2N_{TP}}{2N_{TP}+N_{FP}+N_{FN}}$

Fig. 3 Training and validation loss curves (smoothed) for M1



(707 slices) were used for testing. The cross-entropy loss in Eq. 6 was minimized by a stochastic gradient descent with momentum (SGDM) optimizer. The momentum was set to 0.9 to reduce the oscillations of the weights and biases along the optimization path. Initially, the weights of the encoder were initialized using a pretrained VGG16 [36]. Typically, initializing the network weights with the weights of a pre-trained architecture facilitates a faster convergence to the global optima by leveraging the learned features, particularly at the shallower layers. Conversely, for the decoder weights, we used the MSRA initialization [15], where the weights are initialized from a zero-mean Gaussian distribution whose standard-deviation $\sigma = \sqrt{2/k_l^2 d_l}$, where d_l is the number of filters in layer l , and k_l is the filter size in the same layer. The MSRA initialization have shown to be more efficient in deep networks, preventing to some extent, the problem of diminishing gradient [15]. We set a constant learning rate of 0.001 and a mini-batch size of 2. It is also worth noticing that it is possible to assign a higher learning rate for the decoder arm, and reduce the learning rates associated with encoder layers as to only fine tune the VGG16 pretrained weights. Although it may inexpensively converge faster, this approach do not guarantee the best performance. That is, the decoder weights are essentially coupled with the encoder weights, as they are both sequentially optimized by the same objective function. The training was run for a maximum of 100 epochs. The training and validation loss curves are reported in Fig. 3. To assess

the benefit of our 2D multi-channel approach, we trained the same network using gray-level slices. Each slice was replicated three times in order to fit in the input RGB channels. Other parameters were kept the same for comparison purposes. Throughout this paper, we refer to this method as M1, while we refer to our 2D multi-channel method as M2.

On the other hand, as in most of the medical imaging CAD systems, a leave-one-patient-out cross validation was carried out as well for benchmarking. Particularly, we ran the training and testing algorithms 19 times and averaged the resulting AUC over all the patients. Nevertheless, to cope with the limited computing resources and to reduce the training time, the maximum number of epochs was set to 5.

Results

Benefit of the 2D Multi-Channel Approach

We evaluated the performance of the lesion segmentation task of each class using the metrics described above. These metrics were calculated for all prostate-contained images. The average of each metric is presented in Table 3. Notably, with respect to all metrics, the segmentation of cancerous lesions is significantly enhanced by exploiting 3D contextual information. Also, the mean BF score is significantly improved when using our approach (M2). The performance in terms of IoU seems to be the lowest compared to other metrics, as this metric penalizes the classifier

Table 3 Results of multi-class segmentation on prostate-contained slices

	Accuracy		IoU		Mean BF score		Recall	
	M1	M2	M1	M2	M1	M2	M1	M2
CG	0.908	0.903	0.673	0.657	0.783	0.799	0.815	0.836
PZ	0.920	0.902	0.563	0.599	0.825	0.855	0.901	0.886
Non-prostate	0.998	0.998	0.998	0.998	0.974	0.980	0.980	0.986
CaP	0.876	0.894	0.677	0.679	0.879	0.891	0.916	0.928

Italic emphasis are used to highlight higher performance between M1 and M2

Table 4 Comparison of CPM of CaP results with the literature

	Lemaitre et al. [21]	Wang and Zwiggelaar [46]	Kiraly et al. [16]	Kohl et al. [17]	Yang et al. [50]	M1	M2
Average AUC	0.836	0.883	0.834	—	—	0.9467	0.9569
Recall	—	—	—	0.55	0.844	0.916	0.928
Accuracy	—	—	—	—	0.720	0.876	0.894

on both FP and FN, yet does not take into account TNs. Notably, all metrics are biased towards the Non-prostate class. This however, does not mean that the classifier better classifies the background but rather, the metrics are greatly affected by the relatively large number of background voxels.

Comparison with State-of-Art

To be able to compare with other systems, we report the qualitative and quantitative results for the tasks of prostate segmentation and CaP detection and localization.

CaP Segmentation We assess and compare the performance of our system for the detection and localization of malignant lesions against other recently proposed systems, as can be realized in Table 4. All the reported measures are on a leave-one-patient out cross validation scheme. Note that better performance could be achieved by increasing the maximum number of epochs. However, this will require much more time and computing resources in the adopted cross validation scheme.

For a fair comparison, three recent learning-based systems including [16, 17, 50] are considered. We can notice that our approach outperforms the more traditional pattern recognition and machine learning approach presented by Lemaitre et al. [21] and Wang and Zwiggelaar [46], by more than 11% and 7% average AUC on the same

benchmarking dataset and using the same leave-one-patient-out cross validation protocol. The proposed architecture also outperforms [16, 17, 50] by a significant margin. Note that Kiraly et al. [16] used similar yet shallower architecture, with only five convolutional blocks in each of the decoder and encoder. Expectedly, the system performance is boosted when utilizing a deeper network. That is, more abstract features are extracted in the middle layers, where they indeed contribute to the overall discrimination ability of the network. Besides, the deeper layers contribute to a larger receptive fields and thus better retrieval of the spacial context. Figure 4 qualitatively compares the heatmaps generated by projecting the activations of the SoftMax layer for the two alternative methods explained above, and the output of the CAD system proposed by Lemaitre et al. [21]. The comparison is carried out on the same slices to clearly visualize the CAD performance in each case. It can be easily seen that better performance is achieved using a deep learning-based approach compared to the standard hand-crafted feature-based learning used in [21]. Also, we can notice that the network was able to segment lesions more accurately by the employment of the 3D sliding window approach (Fig. 4).

For benchmarking purposes, we also compare the performance of the proposed approach with that of Wang and Zwiggelaar [46] as shown in Fig. 5. Clearly, our approach predicts the malignant lesions with a better accuracy and

Fig. 4 Results of prostate cancer detection produced by Lemaitre et al. [21], M1 and M2 from left to right, respectively. White contour shows the prostate boundary segmented by a radiologist, while blue contour is the ground truth of malignant lesions. Note that each row shows the same slice and each column shows the performance of the same CAD system

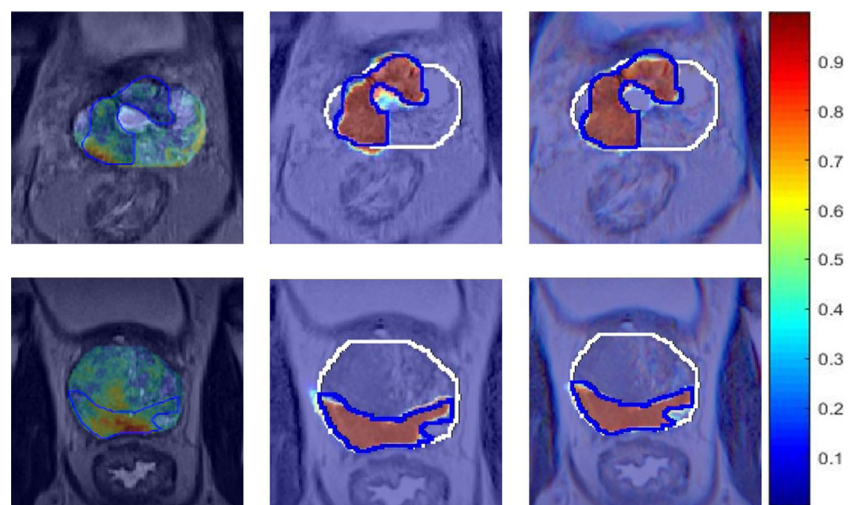


Fig. 5 Results of prostate cancer detection produced by Wang et al. [46] using a 3D texton-based approach are shown in the first row. The second row shows the corresponding performance of our M2 approach on the same slices. Red contours show the ground truth of malignant lesions, while white contours show the boundaries of the peripheral zone segmented by the radiologist

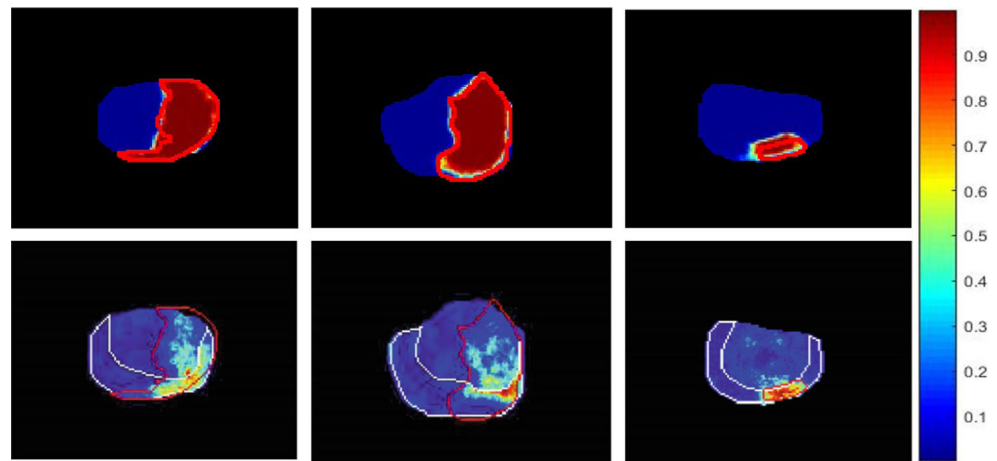
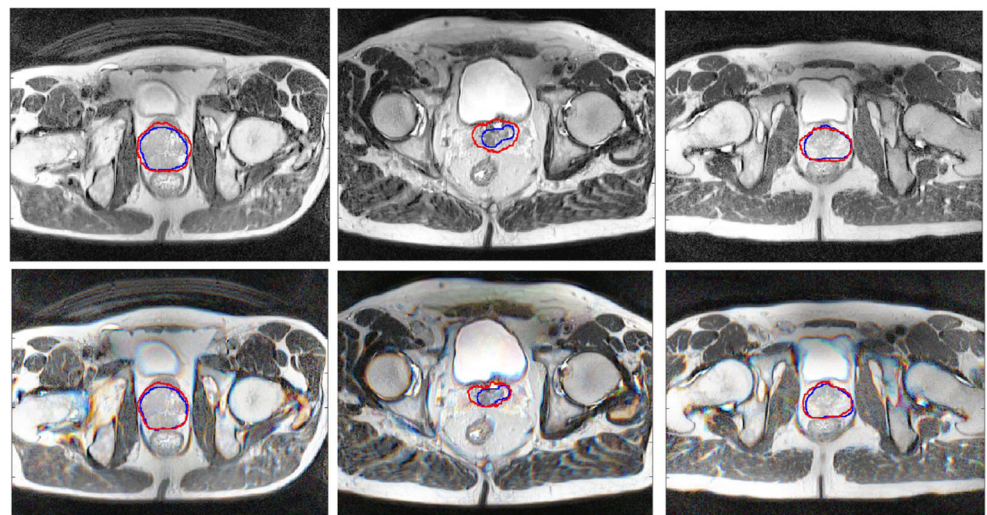


Table 5 Comparison of prostate segmentation results with the literature

	Yu et al. [51]	Drozdzal et al. [6]	Guo et al. [12]	M1	M2
DSC	0.8943	0.874	0.878	0.915	0.921
Precision	—	—	0.916	0.970	0.975

Fig. 6 Prostate segmentation results of M1 and M2. The first row shows examples of segmentation performed using M1. The corresponding segmentation of M2 on the same slices is shown in the second row. The blue contour shows the ground truth segmentation, while the red contour shows the segmentation obtained by the proposed algorithms



less number of false positives as compared to [46]. It is important to point out that Wang and Zwigelaar [46] extracted features for each prostate-contained voxel from a window of size $7 \times 7 \times 3$ which is equivalent to processing three consecutive slices with a receptive field of 7×7 . In our approach, however, the processing takes place on the same depth (i.e., three slices) and with a larger receptive field. In other words, the features extracted by our deep architecture incorporates information from a larger spatial context, which in turn enhances the overall discriminative ability of the system.

Prostate Segmentation A brief comparison between the performance of our approach and other state-of-the-art with respect to the prostate segmentation task is also carried out. Table 5 shows the DSC and the precision of M1, M2 and the prostate segmentation results presented by [6, 12, 51]. Particularly, Yu et al. [51] performed the task using 3D deep CNN with mixed residual connections. Their approach ranked first in the recent PROMIS12 MICCAI challenge. Guo et al. [12] employed a stacked sparse auto-encoder to extract features from MR volumes, which are then used to segment the prostate gland. They validated their model on an in-house dataset of 66 patients. On the other hand, Drozdal et al. [6] tested their 2D CNN-based segmentation algorithm on data from 30 patients. Their model combined an FCN structure which was set to pre-process the input image, along with a ResNet which was used to segment the prostate. Results suggest that our multi-channel approach outperforms other 2D and 3D approaches. More importantly, it outperforms M1 pipeline which uses the same CNN architecture and validation protocol. The improvement of segmentation caused by the exploitation of 3D information suggests that a 2D network was able to learn 3D features with no extra computational cost.

For a qualitative evaluation, Fig. 6 depicts samples of prostate segmentation showcasing the gains achieved by using the proposed approach.

Discussion and Conclusions

In this paper, we proposed a simple yet efficient, deep learning-based approach for joint prostate segmentation and CaP diagnosis on MR images. From our experiments, we draw two general conclusions. First, the incorporation of 3D spatial information through a 2D multi-channel approach is not only possible but also beneficial, and generally applicable to similar medical images with no extra computational cost. This fusion approach allows for the incorporation 3D contextual information in a 2D-based pipeline. Second, the use of a deep convolutional encoder-decoder network for the segmentation of volumetric medical images

yields superior results compared to other state-of-the-art approaches. Also, using the proposed pipeline, the potential of T2W imaging modality was exploited, and the performance of the mono-modal system was comparable to its multi-modal counterparts. Due to the limited availability of large annotated datasets, our work was thus limited. Accordingly, our future work will focus on validating the presented approach on a larger and more diverse dataset to guarantee better generalization.

Acknowledgements The authors would also like to thank Dr. Waleed Hassen and Dr. Eric Vens from Cleveland Clinic, Abu Dhabi, and Dr. Salah El-Rai from Sheikh Khalifa General Hospital for their support and collaboration.

Funding Information This work is support by a research grant from Al-Jalila foundation ref. AJF-201616.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Alkadi R, Taher F, El-Baz A, Naoufel W: Early diagnosis and staging of prostate cancer using magnetic resonance imaging: State of the art and perspectives. In: Prostate cancer imaging: An engineering and clinical perspective, chapter 2. Taylor & Francis, In-press
2. Badrinarayanan V, Kendall A, Cipolla R: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495, 2017
3. Chan I, Wells W, Mulkern RV, Haker S, Zhang J, Zou KH, Maier SE, Tempany C: Detection of prostate cancer by integration of line-scan diffusion, t2-mapping and t2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Med Phys* 30(9):2390–2398, 2003
4. Chawla NV, Bowyer KW, Hall LO, Philip Kegelmeyer W: Smote: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357, 2002
5. Csurka G, Larlus D, Perronnin F, Meylan F: What is a good evaluation measure for semantic segmentation? In: *BMVC*, volume 27, p 2013. Citeseer, 2013
6. Drozdal M, Chartrand G, Vorontsov E, Shakeri M, Di Jorio L, An T, Romero A, Bengio Y, Pal C, Kadoury S: Learning normalized inputs for iterative estimation in medical image segmentation. *Med Image Anal* 44:1–13, 2018
7. Eigen D, Fergus R: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *Proceedings of the IEEE international conference on computer vision*, 2015, pp 2650–2658

8. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F: Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int J Cancer* 136(5):E359–86, 2015
9. Fütterer JJ: Multiparametric mri in the detection of clinically significant prostate cancer. *Korean J Radiol* 18(4):597–606, 2017
10. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J: A review on deep learning techniques applied to semantic segmentation, 2017. arXiv:1704.06857
11. Greenspan H, van Ginneken B, Summers RM: Guest editorial deep learning in medical imaging overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 35(5):1153–1159, 2016
12. Guo Y, Gao Y, Shen D: Deformable mr prostate segmentation via deep feature learning and sparse patch matching. *IEEE Trans Med Imaging* 35(4):1077–1089, 2016
13. Hall MA: Correlation-based feature selection of discrete and numeric class machine learning, 2000
14. Han H, Wang W-Y, Mao B-H: Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: *International Conference on Intelligent Computing*, pp 878–887. Springer, 2005
15. He K, Zhang X, Ren S, Sun J: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*, pp 1026–1034, 2015
16. Kiraly AP, Nader CA, Tuysuzoglu A, Grimm R, Kiefer B, El-Zehiry N, Kamen A: Deep convolutional encoder-decoders for prostate cancer detection and classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp 489–497. Springer, 2017
17. Kohl S, Bonekamp D, Schlemmer H-P, Yaqubi K, Hohenfellner M, Hadaschik B, Radtke J-P, Maier-Hein K: Adversarial networks for the detection of aggressive prostate cancer, 2017. arXiv:1702.08014
18. Kumar D, Wong A, Clausi DA: Lung nodule classification using deep features in ct images. In: *2015 12th conference on computer and robot vision (CRV)*, pp 133–138. IEEE, 2015
19. Lemaitre G: Computer-aided diagnosis for prostate cancer using multi-parametric magnetic resonance imaging. PhD thesis, Ph. D. dissertation, Universitat de Girona and Université de Bourgogne, 2016
20. Lemaître G, Martí R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: A review. *Comput Biol Med* 60:8–31, 2015
21. Lemaitre G, Martí R, Rastgoo M, Mériaudeau F: Computer-aided detection for prostate cancer detection based on multi-parametric magnetic resonance imaging. In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pp 3138–3141. IEEE, 2017
22. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H: Computer-aided detection of prostate cancer in mri. *IEEE Trans Med Imaging* 33(5):1083–1092, 2014
23. Litjens G, Kooi T, Bejnordi BE, Setio A, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI: A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88, 2017
24. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp 3431–3440
25. Lv D, Guo X, Wang X, Zhang J, Fang J: Computerized characterization of prostate cancer by fractal analysis in mr images. *J Magn Reson Imaging* 30(1):161–168, 2009
26. Mani I, Zhang I: knn approach to unbalanced data distributions: a case study involving information extraction. In: *Proceedings of workshop on learning from imbalanced datasets*, vol 126, 2003
27. Mazurowski MA, Buda M, Saha A, Bashir MR: Deep learning in radiology: an overview of the concepts and a survey of the state of the art, 2018. arXiv:1802.08717
28. Puech P, Betrouni N, Makni N, Dewalle A-S, Villers A, Lemaitre L: Computer-assisted diagnosis of prostate cancer using dce-mri data: design, implementation and preliminary results. *Int J Comput. Assist Radiol Surg* 4(1):1–10, 2009
29. Qi CR, Hao SU, Nießner M, Dai A, Yan M, Guibas LJ: Volumetric and multi-view cnns for object classification on 3d data. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp 5648–5656
30. Rampun A, Zheng L, Malcolm P, Tiddeman B, Zwigelaar R: Computer-aided detection of prostate cancer in t2-weighted mri within the peripheral zone. *Phys Med Biol* 61(13):4796, 2016
31. Reda I, Shalaby A, Khalifa F, Elmogy M, Aboufotouh A, El-Ghar MA, Hosseini-Asl E, Werghi N, Keynton R, El-Baz A: Computer-aided diagnostic tool for early detection of prostate cancer. In: *IEEE international conference on image processing (ICIP)*, pp 2668–2672. IEEE, 2016
32. 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
33. Roth HR, Lu Le, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM: Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging* 35(5):1170–1181, 2016
34. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkbey E, Summers RM: A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp 520–527. Springer, 2014
35. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L: ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115(3):211–252, 2015
36. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition, 2014. arXiv:1409.1556
37. Smith MR, Martinez T, Giraud-Carrier C: An instance level analysis of data complexity. *Mach Learn* 95(2):225–256, 2014
38. 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
39. Tiwari P, Rosen M, Madabhushi A: A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (mrs). *Med Phys* 36(9Part1):3927–3939, 2009
40. Tiwari P, Viswanath S, Kurhanewicz J, Sridhar A, Madabhushi A: Multimodal wavelet embedding representation for data combination (mawer): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR Biomed* 25(4):607–619, 2012
41. Trigui R, Mitéran J, Walker PM, Sellami L, Ben Hamida A: Automatic classification and localization of prostate cancer using multi-parametric mri/mrs. *Biomed Signal Process Control* 31:189–198, 2017
42. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC: N4itk: Improved n3 bias correction. *IEEE Trans Med Imaging* 29(6):1310–1320, 2010
43. Viswanath S, Bloch BN, Genega E, Rofsky N, Lenkinski R, Chappelow J, Toth R, Madabhushi A: A comprehensive

- segmentation, registration, and cancer detection scheme on 3 tesla in vivo prostate dce-mri. In: International conference on medical image computing and computer-assisted intervention, pp 662–669. Springer, 2008
44. Viswanath SE, Bloch NB, Chappelow JC, Toth R, Rofsky NM, Genega EM, Lenkinski RE, Madabhushi A: Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo t2-weighted mr imagery. *J Magn Reson Imaging* 36(1):213–224, 2012
 45. Vos PC, Barentsz JO, Karssemeijer N, Huisman HJ: Automatic computer-aided detection of prostate cancer based on multi-parametric magnetic resonance image analysis. *Phys Med Biol* 57(6):1527, 2012
 46. Wang L, Zwiggelaar R: 3d texton based prostate cancer detection using multiparametric magnetic resonance imaging. In: Annual conference on medical image understanding and analysis, pp 309–319. Springer, 2017
 47. Wang Z, Liu C, Cheng D, Wanga L, Yang X, Chengb K-TT: 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*, 2018
 48. Zhirong WU, Song S, Khosla A, Fisher YU, Zhang L, Tang X, Xiao J: 3d shapenets; A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp 1912–1920
 49. Yang X, Liu C, Wang Z, Yang J, Min HL, Wang L, Cheng K-TT: Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric mri. *Med Image Anal* 42:212–227, 2017
 50. Yang X, Wang Z, Liu C, Le HM, Chen J, Cheng K-TT, Wang L: Joint detection and diagnosis of prostate cancer in multi-parametric mri based on multimodal convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 426–434. Springer, 2017
 51. Lequan YU, Yang X, Chen H, Qin J, Heng P-A: Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In: AAAI, 2017, pp 66–72
 52. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A: Learning deep features for discriminative localization. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2921–2929. IEEE, 2016