RIT®

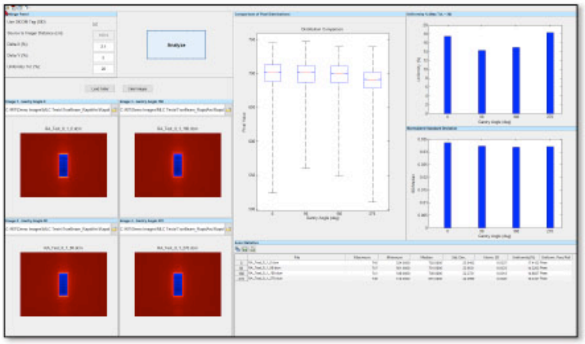
EXPLORE WHAT'S NEW IN

THE RIT FAMILY OF PRODUCTS VERSION 6.11

MEDICAL PHYSICS' MOST ADVANCED QA SOFTWARE

Varian Halcyon® RapidArc® & Picket Fence Analysis

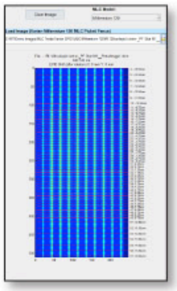
By popular demand, RIT has extended our traditional RapidArc MLC analyses to support Varian Halcyon LINACs. In addition to the Millennium 120 and HD120 models, Halcyon users now have the option to perform fully automated RapidArc QA, with the flexibility to use distal, proximal, or dual MLC configurations.



New Routines & Workflow for Varian RapidArc® Tests

The new RapidArc Test 0.1 (dMLC Dosimetry) interface simultaneously loads and analyzes four cardinal gantry angle images and provides full automation for quick results. The software also features the new RapidArc Test 2 (Dose Rate and Gantry Speed) and Test 3 (Leaf Speed) routines for analyzing the various test modes against an open field image.

Enhanced MLC QA Automation Workflow



RapidArc Tests 0.2, 1.1, and 1.2, and the standard Picket Fence tests have been enhanced with (1) automated pre-processing for rotational and translation alignment, (2) expanded individual leaf analysis, and (3) support for a variety of common EPID sizes: 30x30, 30x40, 40x30, 40x40, and 43x43 cm.

Plan-Based Calibration for TomoTherapy® Registration

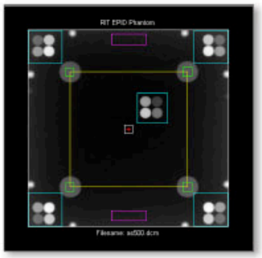
Utilize RIT's patented PBC technology to simplify your TomoTherapy Registration process for patient QA. A calibration file is no longer needed, scanner warm up is minimized, and variation in development time after exposure is no longer a significant factor. This saves you time and money, and potentially improves patient throughput.

Enhanced & Expanded Isocenter Optimization

RIT's popular 3D Winston-Lutz (Isocenter Optimization) routine now supports analysis on Elekta Unity machines. The updated interface will also display a range of 3 to 16 images within the same interface window. File names for exported results are automatically generated, saving users time. Choose to pass or fail on the Maximum Machine Deviation (R) or the Total Maximum Deviation (W), which includes the ball setup error.

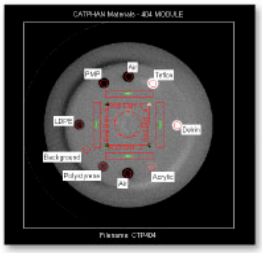
EPID QA Phantom Couch Analysis

RIT software will analyze the EPID QA phantom when placed directly on the couch, providing an analysis closer to realistic patient positioning. The phantom's magnified dimensions are automatically calculated from the phantom size. This update provides physicists with more setup options when performing phantom analyses.



Expanded Phantom Analyses

RIT's Catphan Materials analysis will now analyze both horizontal slices and both vertical slices with ramps, in addition to an added measure of the yaw and tilt of the phantom, based on these ramps. Analyses for the QC-3 and QC-kV1 phantoms now feature a uniformity analysis.



CLICK TO EXPLORE THE NEW FEATURES IN DETAIL

RADIMAGE.COM

+1.719.590.1077, Opt. 4 // sales@radimage.com // Connect with RIT on social media @RIT4QA

Prostate Cancer Segmentation from MRI by a Multistream Fusion Encoder

Mingjie Jiang[1,*], Baohua Yuan[1,2,*], Weixuan Kou[1], Wen Yan[1,4], Harry Marshall[3], Qianye Yang[4], Tom Syer[5], Shonit Punwani[5], Mark Emberton[6], Dean C. Barratt[4], Carmen C. M. Cho[7], Yipeng Hu[4], Bernard Chiu[1]

1.Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China
2.Aliyun School of Big Data, Changzhou University, Changzhou, China
3.Schulich School of Medicine & Dentistry, Western University, Ontario, Canada
4.Centre for Medical Image Computing; Wellcome/EPSRC Centre for Interventional & Surgical Sciences; Department of Medical Physics & Biomedical Engineering, University College London, London, U.K.
5.Centre for Medical Imaging, University College London, London, U.K.
6 Division of Surgery & Interventional Science, University College London, London, U.K.
7.Prince of Wales Hospital and Department of Imaging and Intervention Radiology, Chinese University of Hong Kong, Hong Kong SAR, China

Version typeset March 20, 2023

Corresponding author: Bernard Chiu, email: bcychiu@cityu.edu.hk

Abstract

Background: Targeted prostate biopsy guided by multiparametric magnetic resonance imaging (mpMRI) detects more clinically significant lesions than conventional systemic biopsy. Lesion segmentation is required for planning MRI-targeted biopsies. The requirement for integrating image features available in T2-weighted and diffusion-weighted images poses a challenge in prostate lesion segmentation from mpMRI.

Purpose: A flexible and efficient multistream fusion encoder is proposed in this work to facilitate the multiscale fusion of features from multiple imaging streams. A patch-based loss function is introduced to improve the accuracy in segmenting small lesions.

Methods: The proposed multistream encoder fuses features extracted in the three imaging streams at each layer of the network, thereby allowing improved feature maps to propagate downstream and benefit segmentation performance. The fusion is achieved through a spatial attention map generated by optimally weighting the contribution of the convolution outputs from each stream. This design provides flexibility for the network to highlight image modalities according to their relative influence on

[*] Equal contribution

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.16374

the segmentation performance. The encoder also performs multiscale integration by highlighting the input feature maps (low-level features) with the spatial attention maps generated from convolution outputs (high-level features). The Dice similarity coefficient (DSC), serving as a cost function, is less sensitive to incorrect segmentation for small lesions. We address this issue by introducing a patch-based loss function that provides an average of the DSCs obtained from local image patches. This local average DSC is equally sensitive to large and small lesions, as the patch-based DSCs associated with small and large lesions have equal weights in this average DSC.

Results: The framework was evaluated in 931 sets of images acquired in several clinical studies at two centers in Hong Kong and the United Kingdom. In particular, the training, validation and test sets contain 615, 144 and 172 sets of images, respectively. The proposed framework outperformed single-stream networks and three recently proposed multistream networks, attaining F_1 scores of 82.2% and 87.6% in the lesion and patient levels, respectively. The average inference time for an axial image was 11.8 ms.

Conclusion: The accuracy and efficiency afforded by the proposed framework would accelerate the MRI interpretation workflow of MRI-targeted biopsy and focal therapies.

I. Introduction

Prostate cancer is the most common non-skin cancer for males in the United States[1]. The first-line screening involves digital rectal examination (DRE) and serum prostate-specific antigen (PSA) tests. A transrectal ultrasound-guided (TRUS-guided) systematic biopsy is recommended if either test indicates a suspicion for prostate cancer[2]. However, TRUS biopsy is non-targeted, resulting in low sensitivity for detecting clinically significant prostate lesions. Targeted biopsy directed by multiparametric MRI (mpMRI) has been shown to be superior to conventional systematic biopsy by increasing the detection of clinically significant cancer while reducing the detection of clinically insignificant cancers, reducing the number of unnecessary biopsies, and reducing the number of biopsy cores required to make a diagnosis[3,4].

The Prostate Imaging Reporting and Data System (PIRADS) is the consensus guideline developed by radiologists for interpreting and reporting findings in mpMRI. Although the PIRADS Version 2 recommended the use of the T2-weighted (T2W), diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) sequence for localization and detection of prostate lesions, DCE imaging only plays a minor role in assessing the clinical significance of peripheral zone lesions when they are equivocally suspected by DWI[5]. Since the establishment of PIRADS Version 2, several investigations reported that biparametric MRI (bpMRI), involving only T2W and DWI, has similar diagnostic accuracy compared to mpMRI[6,7]. The acquisition time required by bpMRI is 17 minutes compared to 45 minutes required for mpMRI[7]. Besides, a physician is required to monitor the potential of allergic reactions to intravenous contrast agents required for DCE imaging, thereby increasing the imaging cost. With the substantial saving of imaging time and cost, bpMRI is a strong alternative to mpMRI. In this paper, we evaluated the prostate lesion segmentation performance of our proposed networks in the bpMRI setting.

Lesion segmentation is useful for MRI-targeted biopsy. Beyond its role in the biopsy workflow, lesion segmentation is also required for any form of focal prostate cancer therapies, such as high-intensity focused ultrasound, cryotherapy, or brachytherapy, which aim to treat localized prostate cancer while not exposing the patients to risks associated with aggressive treatments[8]. Although lesion segmentation can be performed manually, manual segmentation is laborious and prone to observer variability.

I.A. Prior work

The segmentation methods proposed in a few earlier works[9,10,11,12,13] were developed before the publication of the mpMRI consensus guideline on prostate mpMRI acquisition and interpretation[5,14] and were evaluated in samples with quantitative T2 maps available. Although quantitative T2 maps are superior to T2W imaging in that they are not affected by variabilities in repetition time (TR) and bias field inhomogeneity, repeated T2W acquisitions were required at tens of echo times, thereby substantially lengthening the acquisition time. Besides, these algorithms[9,10,11,12,13] were evaluated on a single 2D axial image of each prostate chosen at the mid-gland position with lesions segmented from the peripheral zone only.

Deep learning methods have achieved tremendous success in medical image segmentation, and convolutional neural networks (CNN) have been developed to segment lesions from the entire prostates from images acquired according to the consensus guideline. A major challenge for segmenting lesions from MRI is the development of a framework capable of integrating information available from T2W and diffusion-weighted (DW) images. DW images are analyzed as either one or two image modalities. Some methods examined only the apparent diffusion coefficient (ADC) map, which characterizes the amount of water diffusion computed from DW images with different b values, whereas others involve a high-b DW image (DWI_{hb}) in addition to the ADC map. The DWI_{hb} image provides better visualization of clinically significant cancers in regions adjacent to the anterior fibromuscular stroma and at the apex and base of the prostate[5] and is important to be included. Most CNN-based lesion segmentation methods took an early-fusion approach, in which images acquired with different sequences were stacked together and input to the network as a multi-channel tensor[15,16,17,18,19]. Kohl et al.[15] used a U-Net with an adversarial loss to segment lesions. De Vente et al.[16] used U-Net with soft-label output that simultaneously segmented lesions and determined the Gleason Grade Group[20] on a pixel-by-pixel basis. Schelb et al.[17] performed prostate lesion segmentation using a modified U-Net. Netzer et al.[18] evaluated the effects of increased and diversified training data on prostate lesion segmentation performance using U-Net. Sumathipala et al.[19] proposed a Holistically Nested Edge Detector (HED) network, in which side-outputs were generated at different convolution layers and fused to generate the final segmentation mask.

Late fusion is an alternative paradigm that uses multiple streams to process each imaging modality independently and the high-level features extracted by individual streams are then fused together (typically by concatenation). This strategy was shown to have higher segmentation accuracy than early fusion in multimodality MR segmentation applications[21,22,23]. The late fusion strategy was employed in several lesion segmentation methods from bpMRI. Yang et al.[24] developed a multistream architecture to localize prostate lesions. Lesion localization was weakly supervised by a binary image-based tag indicating the presence/absence of lesion(s). Features were independently extracted from T2W and ADC images to generate an activation map for each stream, with the ADC activation map used as the lesion localization result. This method focuses more on determining a rough location of the lesion and not on providing an accurate segmentation for prostate lesions. Chen et al.[25] proposed a multi-branch U-Net (MB-UNet) which integrated the feature from T2W, ADC and DWI$_{hb}$ modalities using channel concatenation and convolution operations. Seetharaman et al.[26] proposed a two-stream model, Stanford Prostate Cancer Network (SPCNet), to perform prostate cancer segmentation from T2W and ADC images. SPCNet concatenated upsampled features extracted at different depths and applied a final classifier to obtain the final segmentation results. In these late fusion strategies, there was no interaction between different branches before the output features of individual branches were concatenated. The lack of interactions does not allow the difference in the feature maps between different streams to be properly adjusted. Recent multistream architectures allow interactions on a layer-by-layer basis. HyperDenseNet[21] concatenates feature maps extracted at all previous layers in all streams. Although providing connection among all streams, the dense concatenation architecture is not scalable to more streams and larger images. FuseUNet[27] concatenated the feature maps extracted in a master and an assistant stream. The network minimized the attention maps generated in the master and assistant streams at each layer, thereby allowing adjustments of difference in the two streams through knowledge distillation. However, FuseUNet is not symmetric. The choice of the master and assistant modalities would have an effect on the segmentation result. This issue was addressed by the cross-modal self-attention distillation network (CSADNet)[28]. CSADNet has two streams to encode the T2W and ADC images for prostate lesion segmentation. The two encoders interact by an attention distillation mechanism, in which the differences between attention maps generated in the two streams were minimized. Unlike FuseUNet, the loss function evaluating the difference between two attention maps was made symmetric by summing the forward and reverse

Kullback-Leibler divergences. The feature maps encoded by the two streams were fused by first multiplying with their correlation matrix; the two resulting feature maps were then concatenated and decoded. Although shown to have high performance in segmenting prostate lesions, CSADNet, like FuseUNet, compares the difference between the attention maps in a pairwise manner. As such, there is no natural extension to accommodate more than two streams. CSADNet was further constrained to be pairwise by the requirement of the correlation matrix of the feature maps from the two streams at the fusion stage. Although multiple pairwise cost functions and multiple correlation matrices can be used to account for an exhaustive permutation of pairwise comparisons, the computation is more complex, and pairwise comparisons may not be able to account for higher-order correlations. This would be a limitation for prostate lesion segmentation from bpMRI as the clinical guideline recommends consideration of DWI$_{hb}$ in addition to T2W and ADC images[5].

I.B. Contributions

In this paper, we propose a multistream encoding network that allows communication among the T2W, ADC and DWI$_{hb}$ branches at each layer. Propagation of fused feature maps obtained through interactions among streams at each layer improves segmentation performance. The specific contributions of this work are summarized below:

1) We propose a layer-by-layer encoder that integrates feature maps in the T2W, ADC and DWI$_{hb}$ streams in a multiscale manner. Multistream interaction is achieved through a spatial attention map generated by integrating the convolution outputs (high-level features) in the three streams adaptively with the relative contribution of each stream optimized by backpropagation. Multiscale integration is achieved on two levels: first, through highlighting the inputs to the three streams (low-level features) by the spatial attention map generated from high-level features and second, by summing the convolution outputs and the spatially highlighted input features.

2) The Dice similarity coefficient (DSC) penalizes incorrect segmentation of larger regions more than that of smaller regions. As a result, more small regions are incorrectly segmented, thereby lowering segmentation accuracy. We propose a patch-based loss function that equally weights incorrect segmentation of larger and smaller regions in each image patch. We showed that adding patch-based loss improves lesion segmentation performance.

3) We performed extensive experiments on a set of diverse bpMRI data acquired from several clinical studies in multiple medical centers. We demonstrated that the fusion module and the new cost function contributed independently to the improvement of lesion segmentation performance and showed that the proposed architecture has a higher segmentation accuracy than existing single-stream and multistream architectures.

II. Materials and methods

II.A. MR imaging

MRI datasets acquired at two centers are involved in this study. Scanning parameters are summarized in Table ??. The first dataset was acquired from 58 subjects at the Princes of Wales Hospital (PWH), Hong Kong. Research ethics approval has been obtained from the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (2021.003) and the Human Subjects Sub-committee in the City University of Hong Kong (10-2021-22-E). T2W and DW images were acquired according to standards that have been set by the consensus guideline[29]. ADC images were generated using the console available in the scanner from DW images acquired with multiple b-values. Two radiologists with six-year experience categorized regions according to PIRADS detection guideline (Version 2)[5] with consensus reached and regions with PIRADS score of at least 3 were then manually segmented by one of the radiologists.

The second dataset was acquired from 637 prostate cancer patients involved in five clinical trials conducted at the University College London (UCL) Hospital as described in the following publications[30,31,32,33]. The goals of the clinical trials include the assessment of diagnostic accuracy of MRI, compared with TRUS biopsy, with and without previous biopsy[31,32], the comparison of different approaches for registering MRI and TRUS during biopsy[30] and the investigation of the roles of MRI in focal ablation[33]. These studies were approved by the local research ethics committees: SmartTarget Biopsy (14/LO/0830), PROMIS (11/LO/0185), INDEX (NCT01194648), PICTURE (11/LO/1657), and SmartTarget Therapy (14/LO/1375). In total, 873 sets of T2W and ADC and DWI$_{hb}$ images were available as multiple scans were performed for some patients. It is important to note that image partitioning to the training, validation and testing sets in our experiments was on a patient basis (i.e., multiple scans from the same patient belong to one partition). Radiologist contours were obtained for all lesions with Likert scores greater than or equal to 3 and served as the ground truth segmentation for radiologist segmentation. The Likert scheme was based on the outputs of a consensus group[29] convened before the publication of PIRADS version 1[14]. The Likert and PIRADS (version 1) scoring schemes were subsequently found to yield similar results[34,35].

II.B. Preprocessing of biparametric MR images

SimpleITK was used to adjust the relative displacement between the T2W, ADC and DWI$_{hb}$ images[36]. Prostate segmentation was done using a pre-trained CNN model[37]. The bounding box of the prostate boundary was expanded by 25% to form a region of interest (ROI). The ROI was cropped for subsequent lesion segmentation. Since CNN requires input images to be of the same size in each training batch, the ROI in each axial image was resampled to a fixed size of 128×128. Although the size of the images in different training batches may vary in fully convolutional networks, the ROIs of all training images were resampled to a fixed size so that the receptive field of the network has similar fractional coverage of all prostates. The ROI of validation and testing images were also resampled to the same size to match the fractional coverage by the receptive field of the trained model.

II.C. MSFusion-UNet for lesion segmentation

II.C.1. Architecture

Fig. 1 shows the architecture of the proposed multistream fusion U-Net (MSFusion-UNet), in which the proposed fusion encoder is embedded in the U-Net structure to form a multistream image segmentation network. The input and output images are represented by rectangles in Fig. 1 with the number of channels labeled. MSFusion-UNet consists of the T2W, ADC and DWI$_{hb}$ streams. Features extracted in these three streams were fused at each layer along the encoder path through the fusion encoder block, described in detail in Sec. II.C.2. Since the input images are smaller than those evaluated by the conventional U-Net[38] by a factor of 4, the number of filters used in the first layer was reduced to 16, as compared to 64 in the original U-Net. The outputs of the encoder paths in the T2W, ADC and DWI$_{hb}$ streams were decoded independently using the U-Net decoder shown in Fig. 1. At the end of the decoder path, the output feature maps from the three streams were concatenated. A pixel classification layer was used to generate a lesion probability map from the concatenated feature maps. The final segmentation map was generated by binarizing the output map of the classifier using a threshold of 0.5.

II.C.2. Fusion encoder module

Fig. 2 shows the encoder developed to fuse features extracted in the T2W, ADC and DWI$_{hb}$ streams. The proposed fusion encoder integrates features extracted in the three streams in a multiscale manner. The inputs from the T2W, ADC and DWI$_{hb}$ streams are denoted by T, A, D, respectively. Features maps, denoted by $F(T)$, $F(A)$ and $F(D)$, were independently generated by the three streams using the double convolutional blocks (DCB) shown in Fig. 1. The input and the filtered feature maps can be considered as low-level and high-level feature maps. A fusion map F_{map} was generated by combining the high-level feature maps:

$$F_{\text{map}} = \sigma(F(T) \circledast k_{1\times1}^{(1)}) + \sigma(F(A) \circledast k_{1\times1}^{(1)}) + \sigma(F(D) \circledast k_{1\times1}^{(1)}), \tag{1}$$

where σ is the sigmoid function. $X \circledast k_{1\times1}^{(1)}$ denotes the convolution of the feature map X with an 1×1 convolution with one kernel. F_{map} was then used as a spatial attention map to highlight the low-level feature map (i.e., A, T and D), as described in Eq. 2. To facilitate the mathematical expression for highlighting the n-channel feature maps A, T and D using a one-channel spatial attention map, we define F_{map}^{rep} as an n-channel map that replicates F_{map} to match the size of A, T and D and express the outputs as element-wise multiplication between the feature maps and F_{map}^{rep}. With this definition, the outputs in the three streams, labeled y_A, y_T and y_D in Fig. 2, can be expressed by:

$$\begin{aligned} y_A &= F(A) + \alpha A \cdot F_{map}^{rep} \\ y_T &= F(T) + \beta T \cdot F_{map}^{rep} \\ y_D &= F(D) + \gamma D \cdot F_{map}^{rep} \end{aligned} \tag{2}$$

where $\cdot$ represents element-wise multiplication and α, β and γ are the learnable coefficients representing the relative weights of F_{map} and they sum up to unity. We note that the outputs y_A, y_T and y_D are the sum of high-level features and spatially highlighted low-level features.

II.C.3. Loss function

Lesion segmentation is a pixel-by-pixel classification task involving a highly unbalanced data set, as the number of pixels with lesions is much smaller than that without lesions. Compared to the mean-square-error and binary cross-entropy, the Dice loss, as a metric measuring the relative overlap between the algorithm and manual segmentation, is more suitable for this application and was used for optimizing the proposed segmentation framework. The following loss is referred to as the global DSC to differentiate it from the patch-based DSC described later:

$$\mathcal{L}_{global} = \mathcal{L}_{DSC}(p, \widehat{p}) = 1 - \frac{2\sum_{i=1}^{N} p_i \widehat{p}_i}{\sum_{i=1}^{N}(p_i + \widehat{p}_i)} \tag{3}$$

where p is a binary map indicating whether each pixel of the input image is within a manually segmented lesion, whereas $\widehat{p}$ is the map representing the probability of each pixel being inside a lesion, as determined by the algorithm. These two maps have the same size as the input image and consist of N pixels, denoted by $\{p_i\}_{i=1}^{N}$ and $\{\widehat{p}_i\}_{i=1}^{N}$.

$\mathcal{L}_{global}$ defined in Eq. 3 is more sensitive to incorrect segmentation for large lesions than small lesions. However, prostate lesions are multi-focal and some of them are small. To improve segmentation performance for small lesions, we propose a patch-based loss function that puts less emphasis on the relative size of lesions. The loss function zooms into each patch of size $S_p \times S_p$ to investigate how well lesions are detected within the patch. The patch-based loss function $\mathcal{L}_{patch}$ consists of two components as quantified below:

$$\mathcal{L}_{patch} = \mathcal{L}_{DSC}(f^{max}_{S_p \times S_p}(p), f^{max}_{S_p \times S_p}(\widehat{p})) + \frac{1}{n_p}\sum_{k=1}^{n_p} \omega_k \mathcal{L}_{DSC}(p^{(k)}_{S_p \times S_p}, \widehat{p}^{(k)}_{S_p \times S_p}) \tag{4}$$

The first term of $\mathcal{L}_{patch}$ involves the maxpool operation $f^{max}_{S_p \times S_p}$, which reduces the dimensions of the 128×128 output images to $\frac{128}{S_p} \times \frac{128}{S_p}$. Each "pixel" in the resulting $\frac{128}{S_p} \times \frac{128}{S_p}$ image shows the maximum value in a $S_p \times S_p$ patch in the algorithm segmented mask $\widehat{p}$ or the manually segmented mask p. The DSC of the maxpooled image penalizes incorrect segmentation in each patch equally, regardless of whether the incorrect segmentation inside the patch was for a large or small lesion. The second term takes a *patch-based* average of the DSC obtained in each $S_p \times S_p$ patch. $p^{(k)}_{S_p \times S_p}$ and $\widehat{p}^{(k)}_{S_p \times S_p}$ denote the k^{th} patches of size $S_p \times S_p$ generated by uniformly splitting p and $\widehat{p}$, respectively, whereas n_p is the total number of patches, which is $(\frac{128}{S_p})^2$ in our studies. In contrast with the $\mathcal{L}_{global}$, this locally averaged DSC weights the DSC in each involved patch equally, whereas the global DSC weights overlap according to the lesion size. w_k is a binary weight that selects patches with a maximum p or $\widehat{p}$ of at least 0.5 to be involved in the DSC calculation [i.e., this DSC only involves patches with a valid manual ($p = 1$) or algorithm ($\widehat{p} \geq 0.5$) segmentation]. The proposed network was trained to optimize the total loss:

$$\mathcal{L}_{total} = \mathcal{L}_{global} + c\mathcal{L}_{patch}, \tag{5}$$

where the weight c was optimized according to Sec. III.B.1.

III. Experiments

III.A. Evaluation metrics and statistical analyses

III.A.1. Area-based metrics

The lesion segmentation performance were evaluated by the Dice similarity coefficient (DSC) and sensitivity (SEN) as defined below:

$$DSC = \frac{2\,|A \cap M|}{|A| + |M|},\ SEN = \frac{|A \cap M|}{|M|}, \tag{6}$$

where A and M are the algorithm-generated and the manual segmentation masks, respectively. $|\cdot|$ measures the area of region. In the evaluation, we measured a DSC and an SEN for each slice and then averaged the measurements for all slices involved in testing.

III.A.2. Lesion-level metrics

Lesion localization performance was assessed by lesion-level precision and recall, the computation of which requires the definition of true positive (TP), false negative (FN) and false positive (FP). Object detection literature declares TP if a gold-standard boundary has an intersection of union (IOU) with its closest algorithm segmented boundary higher than a preset threshold and FN otherwise[39,40]. FP occurs when an algorithm-segmented boundary does not have an IOU with the closest manually segmented boundary above the preset threshold. The same definition can be made with DSC instead of IOU.

However, we note that because IOU and DSC are symmetric (i.e., the metrics do not depend on which boundary is the gold standard boundary and which is the algorithm-segmented boundary), a one-to-one correspondence between the gold standard and algorithm segmentation boundaries is required to be established before quantification[40]. Fig. 3 shows two cases demonstrating the need for one-to-one correspondence may not allow multi-focal cancers to be assessed properly. In Fig. 3a, one of the gold-standard boundaries would be considered as FN (thereby giving a recall of 0.5) although it is completely covered by the algorithm boundary, whereas in Fig. 3b, one of the algorithm boundaries would be considered as FP (thereby giving a precision of 0.5) although it is completely inside the gold-standard boundary.

Yan et al.[41] proposed a set of asymmetric recall and precision metrics to address the issue, which are used for evaluation in this study. Each of the I gold standard lesion volumes $\{M^i\}_{i=1}^I$ is considered a TP if the overlap with any J algorithm segmented lesion volumes $\{A^j\}_{j=1}^J$, single or multiple, is larger or equal to a preset threshold τ and FN otherwise. That is, M^i is a true positive if $\sum_{j=1}^J |M^i \cap A^j|/|M^i| \geq \tau$ and FN otherwise. The above definitions of TP and FN allow the calculation of recall, defined by $TP/(TP+FN)$. Similarly, for the calculation of precision, each A^j is considered a TP if the overlap with a single or multiple M^i is greater than or equal to τ and FP otherwise (i.e., A^j is a TP if $\sum_{i=1}^I |M^i \cap A^j|/|A^j| \geq \tau$ and FP otherwise). The above definitions of TP and FP allow the calculation of precision, defined by $TP/(TP+FP)$. As the metrics are lesion-based and lesions are 3D entities, the above overlap was computed on a 3D basis across axial images. Since lesions with volumes less than 0.2cm^3 are considered insignificant according to Epstein's criteria[42], only M^i and A_i with volumes greater than 0.2cm^3 were considered in the assessment. A low threshold of $\tau = 0.1$ was used in this study because substantial inter-observer variability exists in detecting lesions and margins of up to 10 mm are required to be added to MRI boundary for focal treatments[43,44] as lesions are underestimated in MRI even if segmentation is performed manually[45]. A similar threshold was used in McKinney et al.[46], although the threshold was applied on DSC instead of our asymmetric metrics. The F_1 score was calculated to summarize the precision and recall metrics:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{7}$$

III.A.3. Patient-level metrics

The performance quantified by lesion-based metrics is more affected by patients with multiple lesions. It is also of interest to quantify average precision, recall and the F_1 score per patient. Patient-based metrics were computed by averaging the corresponding metrics obtained for individual patients, thereby weighting the segmentation performance on the image of each patient equally.

III.B. Experimental settings

Four sets of experiments were performed as described in the following subsections. Hyperparameter tuning (Sec. III.B.1.) the ablation studies (Sec. III.B.2.) and the comparison with existing methods (Sec. III.B.3.) were performed using the UCL dataset. A total of 873 sets of data in the UCL database were partitioned into training, validation and test sets with 615, 144 and 172 sets of data, respectively. The proposed MSFusion-UNet trained using the UCL dataset was fine-tuned to segment lesions in the PWH dataset, as described in Sec. III.B.4. The purpose of the experiments on the PWH dataset was to evaluate how well the model trained in a separate dataset generalizes to a new dataset acquired from another center.

III.B.1. Hyperparameter tuning

The weight c of $\mathcal{L}_{patch}$ in Eq. 5 and patch size S_p in Eq. 4 were tuned by sequential optimization. When tuning c, S_p was set as 32. When tuning patch size, c had been optimized and was held at the optimized value. The weight c and patch size S_p were optimized based on the average of the area-based DSC and SEN (Eq. 6) on the validation set, denoted by M_{val}. The maximum M_{val} attained throughout the entire training process was used to assess the performance of each setting. This is a local optimum search and the optimized hyperparameters are not guaranteed to be the global optimum.

III.B.2. Ablation studies

The proposed algorithm has four major features: (1) the three images were processed individually by different streams; (2) a fusion encoder was introduced to integrate feature maps generated by the three streams on a layer-by-layer basis; (3) the weights for each stream at each layer were tuned for optimal segmentation performance (i.e., α, β, γ in Eq. 2); (4) development of the patch-based loss function $\mathcal{L}_{patch}$. Ablation studies were performed to evaluate the contribution of each feature. The baseline model is the single-stream U-Net illustrated in Fig. 4a. The second model, referred to as the *MS* model, implements the multistream structure without the fusion module, as illustrated in Fig. 4b. The three images were processed independently. The output feature maps from these three streams were concatenated with the final segmentation result generated by a classifier consisting of a 1×1 convolution followed by sigmoid activation. In the third model, referred to as the *MSFusion* model, the proposed layer-by-layer fusion model was added to the second model, but α, β and γ were fixed at $\frac{1}{3}$. These three parameters were optimized using backpropagation in the fourth model, referred to as the *MSFusion + Tuning* model. The above four models were driven by $\mathcal{L}_{global}$ alone. The fifth model adds $\mathcal{L}_{patch}$ to the loss function, referred to as the *MSFusion + Tuning +* $\mathcal{L}_{patch}$ model.

III.B.3. Comparison with existing methods

We compared the full MSFusion-UNet model (i.e., *MSFusion + Tuning +* $\mathcal{L}_{patch}$) with three widely used segmentation networks: U-Net[38], Residual U-Net (ResU-Net)[47] and Deeplabv3+[48]. In this comparison, these three networks were configured as single-stream models, in which images of three modalities were concatenated and input to a single encoder-decoder combination, as shown in Fig. 4a. In addition, we also compared with recently published multistream models: the multi-branch UNet (MB-UNet)[25], the SPCNet[26] and the CSADNet[28]. All methods have been tested for prostate lesion segmentation. Tukey's multiple comparison tests were performed to compare the DSC and SEN generated by the proposed MSFusion-UNet with those by existing single-stream and multistream networks. In our experiment, the proposed MSFusion-UNet, MB-UNet and SPCNet have three streams (i.e., T2W, ADC and DWI$_{hb}$), whereas CSADNet processed only the T2W and ADC images, as it has a pairwise structure difficult to be extended to support more than two streams, as detailed the introduction. In our evaluation, the best performance of a method is compared with that of another method. As CSADNet cannot be expanded to three streams, its best performance is attained by processing two streams. On the other hand, methods that can accommodate three streams would generate their best performance when three streams are involved. CSAD-Net is limited to two streams and the comparison of the best performance of the two-stream CSADNet with the three-stream networks would duly include the impact of this limitation. MB-UNet and SPCNet were implemented as a 2.5D network involving three neighbouring image slices in the respective papers[25,26]. As a fair comparison focusing on the evaluation of two-dimensional multistream networks, we compared the performance of a two-dimensional version of these networks with our proposed network in our data set.

III.B.4. Evaluation of the trained model in PWH dataset

The full MSFusion-UNet trained using the UCL dataset was directly applied to segment lesions from half of the patient images in the PWH dataset (i.e., 29 images). We refer to this setting as the pre-trained setting. MSFusion-UNet was then fine-tuned using the remaining 29 images in the PWH dataset. The fine-tuned model was used to segment the images previously processed in the pre-trained setting. The performance in this fine-tuned setting was compared with that in the pre-trained setting.

III.C. Implementation

Adam optimization was used in training with a learning rate of 0.001. All CNNs were trained for 100 epochs with a batch size of 32. The training and testing are performed on Ubuntu 16.04 system with 32GB memory, an Intel(R) Core(TM) i7-9700K CPU of 3.60 GHz and an Nvidia GeForce 2080 Ti graphics card of 11GB memory. All methods described in Sec. III.B.3. were trained with the same number of epochs as the proposed method.

III.D. Training set augmentation

Data augmentation was performed by transforming the original images of three modalities simultaneously using the following operations: (a) Flipping: An image was randomly selected to be transformed by one of the following three flipping operations: vertical, horizontal or vertical + horizontal flipping operations. The probability of selecting each operation is 1/3. (b) Rotation: An image was rotated about the image center with an angle within the range of -20° to 20° randomly chosen from a uniform probability distribution. (c) Zoom: An image was randomly zoomed within a range of [0.9,1.1] (d) Translation: An image was translated along the x- and y-axes by distances ranging from 0 to 5 pixels. The x and y-translations were randomly chosen from independent uniform probability distributions. (e) Shear Intensity: An image was fixed on an axis and stretched with a shear intensity of 0.05.

IV. Results

IV.A. Hyperparameter tuning

Our model was trained with $c = 0, 0.25, 0.5, 0.75$ and 1. The corresponding M_{val} are 61.5, 59.6, 61.9, 59.5 and 61.1 with the maximum attained at $c = 0.5$. Therefore, $c = 0.5$ was used in the remaining experiments in this paper. The models trained with different c values were evaluated in the test set with the results reported in Table **??**.

With c held at 0.5, our model was trained with patch sizes S_p of $16, 32$ and 64. The corresponding M_{val} were 60.2, 61.9 and 61.8, respectively, with the maximum attained at $S_p = 32$. Therefore, $S_p = 32$ was used in the remaining experiments in this paper. We evaluated the models trained with different S_p values with test data and the result is reported in Table **??**.

IV.B. Ablation studies

Table **??** shows the results of the ablation study described in Sec. III.B.2. The MS model has a higher DSC, sensitivity and F_1 scores in the lesion and patient levels than the baseline. These four metrics were improved by the fixed coefficient fusion encoder (MSFusion) and further improved by tuning the coefficients in the fusion encoder (MSFusion+Tuning). The introduction of the $\mathcal{L}_{patch}$ loss function contributed to a further improvement in sensitivity and the lesion- and patient-level F_1 scores. Fig. 5 shows examples in which the $\mathcal{L}_{patch}$ loss function helped reduce false positive and false negative segmentations.

The multistream fusion module was used in five scale levels in MSFusion-UNet, as shown in Fig. 1. Fig. 6 shows the mean feature maps in the ADC, T2W and DWI$_{hb}$ stream (i.e., y_A, y_T and y_D, respectively) at each level for (a) MSFusion+Tuning+$\mathcal{L}_{patch}$ and (b) MS settings.

The combined feature map F_{map} for each of the five levels is also shown for Setting a. The feature maps were averaged in the channel dimension to facilitate visualization. Although this lesion was primarily detected in DWI$_{hb}$ and, to a lesser extent, in ADC, the lesion was also highlighted in the T2W feature map starting from Level 2 in the MSFusion+Tuning+$\mathcal{L}_{patch}$ setting due to the inter-stream interaction, whereas the lesion was never highlighted clearly in the T2W feature map in the MS setting. Propagation of more discriminative features results in higher segmentation accuracy.

IV.C. Comparison with existing methods

Table ?? shows the performance metrics for the seven deep learning models described in Sec. III.B.3. The proposed method has the highest performance in all metrics. Among the three multistream methods (i.e., MB-UNet[25], SPCNet[26] and CSADNet[28]), only the CSADNet involves interactions between streams during feature encoding. The performance of CSADNet was lower than MB-UNet because it had access to three imaging modalities (i.e., T2W, ADC and DWI$_{hb}$), whereas CSADNet could only process the T2W and ADC due to the pairwise nature of the model. Tukey's multiple comparison tests show that the area-based sensitivity of the proposed method is significantly higher than the six methods involved in the comparison ($6.66 \times 10^{-14} \leq p \leq 6.71 \times 10^{-14}$), whereas DSC is significantly higher than all methods ($2.94 \times 10^{-6} \leq p \leq 2.05 \times 10^{-2}$), except Deeplabv3+ ($p = 0.18$). Fig. 7 shows example lesions segmented by these models, illustrating that the proposed model has a higher precision (first/second rows) and a higher sensitivity to small lesions (third/fourth rows).

Table ?? shows the inference time of the segmentation models evaluated in Table ??. While improving the lesion segmentation performance as presented above, the inclusion of the fusion encoder in the multistream network involved only a small computational overhead. The CSADNet is the only multistream architecture that allows inter-stream interaction during encoding and the interaction was between two streams. While our proposed MSFusion-UNet allows interactions among three streams, the inference time required was only 65% of that required by CSADNet.

IV.D. Evaluation on the PWH dataset

Table **??** shows the evaluation based on the PWH dataset described in Sec. III.B.4. The DSC, lesion- and patient-level precisions of the MSFusion-UNet pre-trained model dropped by 2%, 22% and 22%, respectively, compared to the results in the UCL dataset, but the sensitivity, lesion- and patient-level recalls increased by 8%, 3% and 2%, respectively. As the sample size of the UCL database is large, the variability in the features across lesions is also large. Therefore, the pre-trained model may have identified a wide range of regions as lesions, resulting in high recall but also identifying more false positives. Fine-tuning by training with the PWH data successfully achieved a better balance between precision and recall, and resulted in higher DSC and F_1 scores at the lesion and patient levels.

We also compared the manual segmentation by a second observer with the ground truth manual segmentation performed by a more experienced radiologist. The second observer is a subspecialist genitourinary radiologist with one year of experience in segmenting regions suggestive of prostate cancer. The DSC and SEN are $66.6 \pm 31.7\%$ and $76.5 \pm 35.6\%$, which are higher than the fine-tuned model, but the SEN is lower than the pre-trained model. We also investigated how adding a margin of 1mm to 10mm to the segmented regions would increase SEN. As MRI was found to underestimate the size of prostate lesions, margins of up to 10 mm would be required to obtain adequate coverage of the lesions for focal therapies[43,44]. Fig. 8 shows that SEN of the fine-tuned model surpassed that of the second radiologist with a margin of 2mm, although the model had a lower SEN with no margin. The SEN for the second radiologist segmentation did not improve beyond a margin of 3mm and was 83.8% with a 10mm margin. In contrast, the SEN of the model increased more with margin added and the sensitivity reached 96.1% with a margin of 10mm. This suggests that the proposed MSFusion-UNet segmentation could identify more lesions with a sufficient margin.

V. Discussion

The development of the proposed multistream fusion strategy was motivated by the need to integrate information available in T2W, ADC and DWI_{hb} images for lesion segmentation. In particular, the development of an optimal way to integrate features from multiple streams would improve the network's ability in identifying important features.

We introduced a fusion encoder integrating features extracted in the T2W, ADC and DWI$_{hb}$ images for lesion segmentation. First, the fusion encoder enables the layer-by-layer fusion of the three image branches in a multiscale manner, thereby allowing improved optimization of the feature maps propagated to the layers downstream. Second, the weight of the fusion map used to construct the output from each stream was adaptively determined by backpropagation for optimal performance. We have shown in our ablation experiments that the use of the fusion module and the weight tuning individually contributed to improving lesion segmentation performance.

An important novelty of the proposed MSFusion-UNet is the introduction of an attention mechanism that allows layer-by-layer interaction among multiple streams. Attention mechanisms have been previously proposed to facilitate image fusion between two streams in FuseUNet[27] and CSADNet[28]. Both networks involve the computation of attention maps in two image streams. Image fusion was achieved by implementing cost functions to promote similarity between the attention maps extracted from different streams in multiple layers. The attention module within FuseUNet was shown to improve the performance in breast lesion segmentation, whereas that implemented in CSADNet improved prostate lesion segmentation. While we also fused features extracted in multiple streams through an attention map, we did not impose constraints on the similarity of features extracted by different streams, thereby providing more flexibility in optimizing lesion localization and segmentation based on the loss function. Different streams identify different salient features (e.g., the lesion in Fig. 6 was highlighted to a greater extent by the ADC and DWI_{hb} streams than the T2W stream); enforcing similarity on features extracted by multiple streams at each layer may not benefit segmentation performance. The proposed MSFusion-UNet further promotes flexibility in multistream fusion by introducing adaptive parameters to control the weight of the attention map to be added to the output from each stream. In addition to its flexibility in the flow of information across image streams, the proposed network also provides architectural flexibility that allows integration of more than two streams, whereas FuseUNet and CSADNet are limited to the fusion of two streams partly due to their requirement in enforcing pairwise similarity of features.

Although area-based metrics such as DSC and sensitivity are widely used in evaluating segmentation performance, it is more important in clinical applications (e.g., targeted biopsy, focal therapies) to evaluate the precision and recall in segmenting lesions. Although lesion-level precision and recall have been defined in object detection applications, these definitions require one-to-one correspondence between objects detected by an algorithm and the ground truth and are susceptible to issues illustrated in Fig. 3. The lesion-level evaluation metrics proposed in Yan et al.[41] address the issue and were used in this paper.

With the lesion-level evaluation, we found that the DSC loss (i.e., $\mathcal{L}_{global}$) penalizes incorrect segmentation of large regions more than small regions, resulting in more inaccuracies in segmenting small regions. We address this issue by introducing a patch-based loss function $\mathcal{L}_{patch}$ (Eq. 4) that weights the DSC loss in all image patches equally, regardless of the size of the detected regions being evaluated. We showed that the introduction of $\mathcal{L}_{patch}$ increased the lesion- and patient-level precision and recall. In addition to its contribution to prostate lesion segmentation, the proposed patch-based loss function would potentially improve the segmentation performance for other multi-focal lesions, such as breast lesion segmentation[46], although a thorough validation is required to verify the benefit in the new application.

We compared the proposed framework with six state-of-the-art single-stream and multistream deep learning segmentation models and demonstrated that the proposed framework performed better in DSC and area-based sensitivity and has higher F_1 scores in the lesion and patient levels than the six existing models. We note that some previous methods were evaluated to have a higher DSC in the original study. For example, Chen et al.[25] reported that MB-UNet has a DSC of 0.63 and an area-based sensitivity of 0.71. In that study, MB-UNet was evaluated with images with PIRADS ≥ 4, whereas our data set comes from different clinical trials and includes lesions less certain to be clinically significant (Likert score ≥ 3). The Likert score was used before PIRADS was established, and the two schemes produce similar clinical results[34,35]. The DSC reported in the current study is higher than most investigations involving deep-learning lesion segmentation approaches. For example, Schelb et al.[17] included lesions with PIRADS ≥ 3 and reported a DSC of 0.35. Several studies involving histopathology-confirmed lesions[15,49,50] reported a DSC ranging from 0.37 to 0.48 and a sensitivity ranging from 0.46 to 0.55. In the current study, our comparisons were conducted with all state-of-the-art methods trained with the same number of epochs and the same training/validation data, and tested with the same testing data, thereby demonstrating in a fair manner that the proposed segmentation framework outperforms other models.

Although our method performed better than existing approaches in a large dataset acquired in multiple centers and in different clinical studies, we acknowledge several limitations of our approach. Our algorithm is slice-based, and therefore, did not account for the continuity exhibited in adjacent axial prostate images as in a 3D CNN. However, although the prostate is depicted by consecutive axial image slices, the continuity of lesions in adjacent axial slices in our image data is limited due to the large slice thickness (3-5mm) in DWI and the size of the lesions, which can be as small as $0.2cm^3$ according to Epstein's criteria[42]. For images with slice thickness substantially larger than the in-plane voxel sizes, there is evidence showing that 3D CNN has a lower performance than 2D CNN[51,52]. Therefore, although thorough evaluation is warranted, 3D or 2.5D processing may not improve the segmentation performance. Moreover, 2D implementation requires less computational power and memory, which provides flexibility for training, inference and further adaptation in the clinical setting. For these reasons, we opted to implement the network in 2D in this study. Second, we recognize that the relative importance of T2W and DWI in detecting prostate lesions depends on whether the lesion is located in the peripheral (PZ) or the transitional zone (TZ)[5]. Therefore, the weights of the three modalities (i.e., α, β, γ) should ideally be trained individually for lesions at PZ and TZ and applied in the testing images according to the lesion position. However, such an approach would necessitate either automated prostate zonal segmentation or a softer delineation of the PZ and TZ in the training and testing images. A thorough evaluation is needed to assess the impact of zonal segmentation accuracy on lesion segmentation accuracy, and a dedicated workflow must be designed to account for cases with multi-focal lesions occurring in both PZ and TZ. Considering that zonal information has been shown to improve prostate lesion segmentation and detection[16,53], we believe that the performance of the proposed network will improve when zonal prior is provided.

VI. Conclusion

A multistream network was proposed to segment prostate lesions from T2W, ADC and DWI_{hb} image modalities. The major innovation of this work is the development of a layer-by-layer encoder allowing multistream and multiscale interactions when encoding image features. The inter-stream interaction was optimized by adaptively weighting the contributions of the three modalities. The introduction of a patch-based DSC loss component in the cost function improves the lesion- and patient-level precision and recall. Extensive experiments involving 931 sets of images acquired in several clinical studies from medical centers in Hong Kong and the United Kingdom show that the proposed network outperformed state-of-the-art single-stream and multistream networks.

Acknowledgements

This work is supported by the Innovation and Technology Commission of Hong Kong (Project No. ITS/387/18) and Research Grant Council of HKSAR, China (Project No. CityU 11205421).

Disclosure

The article is related to a patent filed as a US non-provisional patent titled “Multistream Fusion Encoder for Prostate Cancer Segmentation And Classification” (Inventors: B. Chiu, C.C.M. Cho, M. Jiang and B. Yuan). Mark Emberton receives research support from the United Kingdom’s National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre. He became an NIHR Senior Investigator in 2015. He acts as a consultant/trainer/lecturer to the following companies: Sonacare Inc USA, Profound Medical Inc. Canada, Angiodynamics Inc. USA, NINA Medical Inc. Israel.

References

[1] R. Siegel, K. Miller, A. Jemal, Cancer facts & figures 2016, american cancer society, Atlanta, Georgia (2016) 1–72.

[2] N. Mottet, J. Bellmunt, M. Bolla, E. Briers, M. G. Cumberbatch, M. De Santis, N. Fossati, T. Gross, A. M. Henry, S. Joniau, et al., EAU-ESTRO-SIOG guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent, Eur. Urol. 71 (4) (2017) 618–629.

[3] V. Kasivisvanathan, A. S. Rannikko, M. Borghi, V. Panebianco, L. A. Mynderse, M. H. Vaarala, A. Briganti, L. Budäus, G. Hellawell, R. G. Hindley, et al., MRI-targeted or standard biopsy for prostate-cancer diagnosis, N. Engl. J. Med. 378 (19) (2018) 1767–1777.

[4] M. van der Leest, E. Cornel, B. Israël, R. Hendriks, A. R. Padhani, M. Hoogenboom, P. Zamecnik, D. Bakker, A. Y. Setiasti, J. Veltman, et al., Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naive men with elevated prostate-specific antigen: a large prospective multicenter clinical study, Eur. Urol. 75 (4) (2019) 570–578.

[5] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, et al., PI-RADS prostate imaging–reporting and data system: 2015, version 2, Eur. Urol. 69 (1) (2016) 16–40.

[6] A. Stanzione, M. Imbriaco, S. Cocozza, F. Fusco, G. Rusconi, C. Nappi, V. Mirone, F. Mangiapia, A. Brunetti, A. Ragozzino, et al., Biparametric 3T magnetic resonance imaging for prostatic cancer detection in a biopsy-naïve patient population: a further improvement of PI-RADS v2?, Eur. J. Radiol. 85 (12) (2016) 2269–2274.

[7] K. C. D. Thestrup, V. Logager, I. Baslev, J. M. Møller, R. H. Hansen, H. S. Thomsen, Biparametric versus multiparametric MRI in the diagnosis of prostate cancer, Acta Radiol. Open 5 (8) (2016) 2058460116663046.

[8] M. Valerio, Y. Cerantola, S. E. Eggener, H. Lepor, T. J. Polascik, A. Villers, M. Emberton, New and established technology in focal ablation of the prostate: a systematic review, Eur. Urol. 71 (1) (2017) 17–34.

[9] Y. Artan, M. A. Haider, D. L. Langer, T. H. Van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, J. Trachtenberg, I. S. Yetik, Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields, IEEE Trans. Image Process. 19 (9) (2010) 2444–2455.

[10] Y. Artan, I. S. Yetik, Prostate cancer localization using multiparametric MRI based on

semisupervised techniques with automated seed initialization, IEEE Trans. Inf. Technol. Biomed. 16 (6) (2012) 1313–1323.

[11] D. L. Langer, T. H. Van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson, M. A. Haider, Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI, J. Magn. Reson. Imaging 30 (2) (2009) 327–334.

[12] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, I. S. Yetik, Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class, IEEE Trans. Med. Imag. 28 (6) (2009) 906–915.

[13] S. Ozer, D. L. Langer, X. Liu, M. A. Haider, T. H. Van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, I. S. Yetik, Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI, Med. Phys. 37 (4) (2010) 1873–1883.

[14] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, J. J. Fütterer, ESUR prostate MR guidelines 2012 22 (4) (2012) 746–757.

[15] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, J.-P. Radtke, K. Maier-Hein, Adversarial networks for the detection of aggressive prostate cancer, arXiv preprint arXiv:1702.08014 (2017).

[16] C. de Vente, P. Vos, M. Hosseinzadeh, J. Pluim, M. Veta, Deep learning regression for prostate cancer detection and grading in bi-parametric MRI, IEEE Trans. Biomed. Eng. 68 (2) (2021) 374–383.

[17] P. Schelb, S. Kohl, J. P. Radtke, M. Wiesenfarth, P. Kickingereder, S. Bickelhaupt, T. A. Kuder, A. Stenzinger, M. Hohenfellner, H.-P. Schlemmer, et al., Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment, Radiology 293 (3) (2019) 607–617.

[18] N. Netzer, C. Weißer, P. Schelb, X. Wang, X. Qin, M. Görtz, V. Schütz, J. P. Radtke, T. Hielscher, C. Schwab, et al., Fully automatic deep learning in bi-institutional prostate magnetic resonance imaging: Effects of cohort size and heterogeneity, Invest. Radiol. 56 (12) (2021) 799–808.

[19] Y. Sumathipala, N. S. Lay, B. Turkbey, C. Smith, P. L. Choyke, R. M. Summers, Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks, J. Med. Imaging 5 (4) (2018) 044507.

[20] J. I. Epstein, M. J. Zelefsky, D. D. Sjoberg, J. B. Nelson, L. Egevad, C. Magi-Galluzzi, A. J. Vickers, A. V. Parwani, V. E. Reuter, S. W. Fine, et al., A contemporary prostate cancer grading system: a validated alternative to the Gleason score, Eur. Urol. 69 (3) (2016) 428–435.

[21] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, I. B. Ayed, HyperDense-Net: a hyper-densely connected cnn for multi-modal image segmentation, IEEE Trans. Med. Imag. 38 (5) (2018) 1116–1126.

[22] D. Nie, L. Wang, Y. Gao, D. Shen, Fully convolutional networks for multi-modality isointense infant brain image segmentation, in: Proc. IEEE Int. Symp. Biomed. Imaging, IEEE, 2016, pp. 1342–1345.

[23] A. Pinto, S. Pereira, R. Meier, V. Alves, R. Wiest, C. A. Silva, M. Reyes, Enhancing clinical MRI perfusion maps with data-driven maps of complementary nature for lesion outcome prediction, in: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Springer, 2018, pp. 107–115.

[24] X. Yang, C. Liu, Z. Wang, J. Yang, H. Le Min, L. Wang, K.-T. T. Cheng, Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI, Med. Image Anal. 42 (2017) 212–227.

[25] Y. Chen, L. Xing, L. Yu, H. P. Bagshaw, M. K. Buyyounouski, B. Han, Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch UNet, Med. Phys. 47 (12) (2020) 6421–6429.

[26] A. Seetharaman, I. Bhattacharya, L. C. Chen, C. A. Kunder, W. Shao, S. J. Soerensen, J. B. Wang, N. C. Teslovich, R. E. Fan, P. Ghanouni, et al., Automated detection of aggressive and indolent prostate cancer on magnetic resonance imaging, Med. Phys. (2021).

27 C. Li, H. Sun, Z. Liu, M. Wang, H. Zheng, S. Wang, Learning cross-modal deep representations for multi-modal MR image segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Springer, 2019, pp. 57–65.

28 G. Zhang, X. Shen, Y. Zhang, Y. Luo, J. Luo, D. Zhu, H. Yang, W. Wang, B. Zhao, J. Lu, Cross-modal prostate cancer segmentation via self-attention distillation, IEEE J. Biomed. Health Inform. (2021).

29 L. Dickinson, H. U. Ahmed, C. Allen, J. O. Barentsz, B. Carey, J. J. Futterer, S. W. Heijmink, P. J. Hoskin, A. Kirkham, A. R. Padhani, et al., Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting, Eur. Urol. 59 (4) (2011) 477–494.

30 S. Hamid, I. A. Donaldson, Y. Hu, R. Rodell, B. Villarini, E. Bonmati, P. Tranter, S. Punwani, H. S. Sidhu, S. Willis, et al., The SmartTarget biopsy trial: a prospective, within-person randomised, blinded trial comparing the accuracy of visual-registration and magnetic resonance imaging/ultrasound image-fusion targeted biopsies for prostate cancer risk stratification, Eur. Urol. 75 (5) (2019) 733–740.

31 L. A. Simmons, A. Kanthabalan, M. Arya, T. Briggs, D. Barratt, S. C. Charman, A. Freeman, J. Gelister, D. Hawkes, Y. Hu, et al., The PICTURE study: diagnostic accuracy of multiparametric MRI in men requiring a repeat prostate biopsy, Br. J. Cancer 116 (9) (2017) 1159–1165.

32 H. U. Ahmed, A. E.-S. Bosaily, L. C. Brown, R. Gabe, R. Kaplan, M. K. Parmar, Y. Collaco-Moraes, K. Ward, R. G. Hindley, A. Freeman, et al., Diagnostic accuracy of multi-parametric MRI and trus biopsy in prostate cancer (PROMIS): a paired validating confirmatory study, The Lancet 389 (10071) (2017) 815–822.

33 H. U. Ahmed, L. Dickinson, S. Charman, S. Weir, N. McCartan, R. G. Hindley, A. Freeman, A. P. Kirkham, M. Sahu, R. Scott, et al., Focal ablation targeted to the index lesion in multifocal localised prostate cancer: a prospective development study, Eur. Urol. 68 (6) (2015) 927–936.

34 A. B. Rosenkrantz, R. P. Lim, M. Haghighi, M. B. Somberg, J. S. Babb, S. S. Taneja, Comparison of interreader reproducibility of the prostate imaging reporting and data sys-

tem and likert scales for evaluation of multiparametric prostate MRI, Am. J. Roentgenol. 201 (4) (2013) W612–W618.

[35] A. R. Rastinehad, N. Waingankar, B. Turkbey, O. Yaskiv, A. M. Sonstegard, M. Fakhoury, C. A. Olsson, D. N. Siegel, P. L. Choyke, E. Ben-Levi, et al., Comparison of multiparametric MRI scoring systems and the impact on cancer detection in patients undergoing mr us fusion guided prostate biopsies, PloS one 10 (11) (2015) e0143404.

[36] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare, SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research, J. Digit. Imaging 31 (3) (2018) 290–303.

[37] A. Meyer, M. Rakr, D. Schindele, S. Blaschke, M. Schostak, A. Fedorov, C. Hansen, Towards patient-individual PI-RADS v2 sector map: Cnn for automatic segmentation of prostatic zones from T2-weighted MRI, in: Proc. IEEE Int. Symp. Biomed. Imaging, IEEE, 2019, pp. 696–700.

[38] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Springer, 2015, pp. 234–241.

[39] Z. Cai, N. Vasconcelos, Cascade R-CNN: Delving into high quality object detection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 6154–6162.

[40] R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, E. A. da Silva, A comparative analysis of object detection metrics with a companion open-source toolkit, Electronics 10 (3) (2021) 279.

[41] W. Yan, Q. Yang, T. Syer, Z. Min, S. Punwani, M. Emberton, D. C. Barratt, B. Chiu, Y. Hu, The impact of using voxel-level segmentation metrics on evaluating multifocal prostate cancer localisation, in: Proceedings of Applications of Medical Artificial Intelligence (AMAI), Vol. 13540 of Lecture Notes in Computer Science, Springer, 2022, pp. 128–138.

[42] J. I. Epstein, P. C. Walsh, M. Carmichael, C. B. Brendler, Pathologic and clinical findings to predict tumor extent of nonpalpable (stage t1 c) prostate cancer, Jama 271 (5) (1994) 368–374.

[43] E. Gibson, G. S. Bauman, C. Romagnoli, D. W. Cool, M. Bastian-Jordan, Z. Kassam, M. Gaed, M. Moussa, J. A. Gómez, S. E. Pautler, et al., Toward prostate cancer contouring guidelines on magnetic resonance imaging: dominant lesion gross and clinical target volume coverage via accurate histology fusion, Int. J. Radiat. Oncol. Biol. Phys. 96 (1) (2016) 188–196.

[44] J. Le Nobin, A. B. Rosenkrantz, A. Villers, C. Orczyk, F.-M. Deng, J. Melamed, A. Mikheev, H. Rusinek, S. S. Taneja, Image guided focal therapy for magnetic resonance imaging visible prostate cancer: defining a 3-dimensional treatment margin based on magnetic resonance imaging histology co-registration analysis, J. Urol. 194 (2) (2015) 364–370.

[45] J. Le Nobin, C. Orczyk, F.-M. Deng, J. Melamed, H. Rusinek, S. S. Taneja, A. B. Rosenkrantz, Prostate tumour volumes: evaluation of the agreement between magnetic resonance imaging and histology using novel co-registration software, BJU Int. 114 (2014) E105.

[46] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. S. Corrado, A. Darzi, et al., International evaluation of an ai system for breast cancer screening, Nature 577 (7788) (2020) 89–94.

[47] X. Xiao, S. Lian, Z. Luo, S. Li, Weighted Res-Unet for high-quality retina vessel segmentation, in: Int. Conf. Inf. Technol. Med. Education, IEEE, 2018, pp. 327–331.

[48] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proc. Eur. Conf. Comput. Vis., 2018, pp. 801–818.

[49] Z. Dai, E. Carver, C. Liu, J. Lee, A. Feldman, W. Zong, M. Pantelic, M. Elshaikh, N. Wen, Segmentation of the prostatic gland and the intraprostatic lesions on multiparametic magnetic resonance imaging using mask region-based convolutional neural networks, Advances in Radiation Oncology 5 (3) (2020) 473–481.

[50] W. Jung, S. Park, K.-H. Jung, S. I. Hwang, Prostate cancer segmentation using manifold mixup u-net, in: Proceedings of the Medical Imaging with Deep Learning, 2019.

[51] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, K. H. Maier-Hein, Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features, in: International workshop on statistical atlases and computational models of the heart, Springer, 2017, pp. 120–129.

[52] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., nnU-net: Self-adapting framework for U-net-based medical image segmentation, arXiv preprint arXiv:1809.10486 (2018).

[53] M. Hosseinzadeh, P. Brand, H. Huisman, Effect of adding probabilistic zonal prior in deep learning-based prostate cancer detection, in: Int. Conf. Med. Imaging Deep. Learn., London, United Kingdom, 2019.

Figure captions

1. The architecture of the proposed multistream fusion U-Net (MSFusion-UNet). The number below each feature block shows the number of channels in the block. MSFM: multistream fusion module, the detailed structure of which is shown in Fig. 2; Conv3×3, Conv1×1: Convolutions with 3×3 and 1×1 kernels with stride $= 1$, MaxPool: maxpooling with a 2×2 kernel and stride = 2; UpSample: upsample operation with 2×2 upsampling factor; Concat: channel concatenation; BatchNorm: batch normalization; ReLU: rectified linear unit; Sigmoid: sigmoid function.

2. The architecture of the multistream fusion encoder module. DCB: Double convolution block shown in Fig. 2; F_{map} is a one-channel attention map and F_{map}^{rep} is an multi-channel map that replicates F_{map} to match the size of A, T and D.

3. Two cases illustrating issues of object detection metrics requiring one-to-one correspondence of boundaries. The gold standard boundary and the algorithm boundaries were represented in red and blue, respectively.

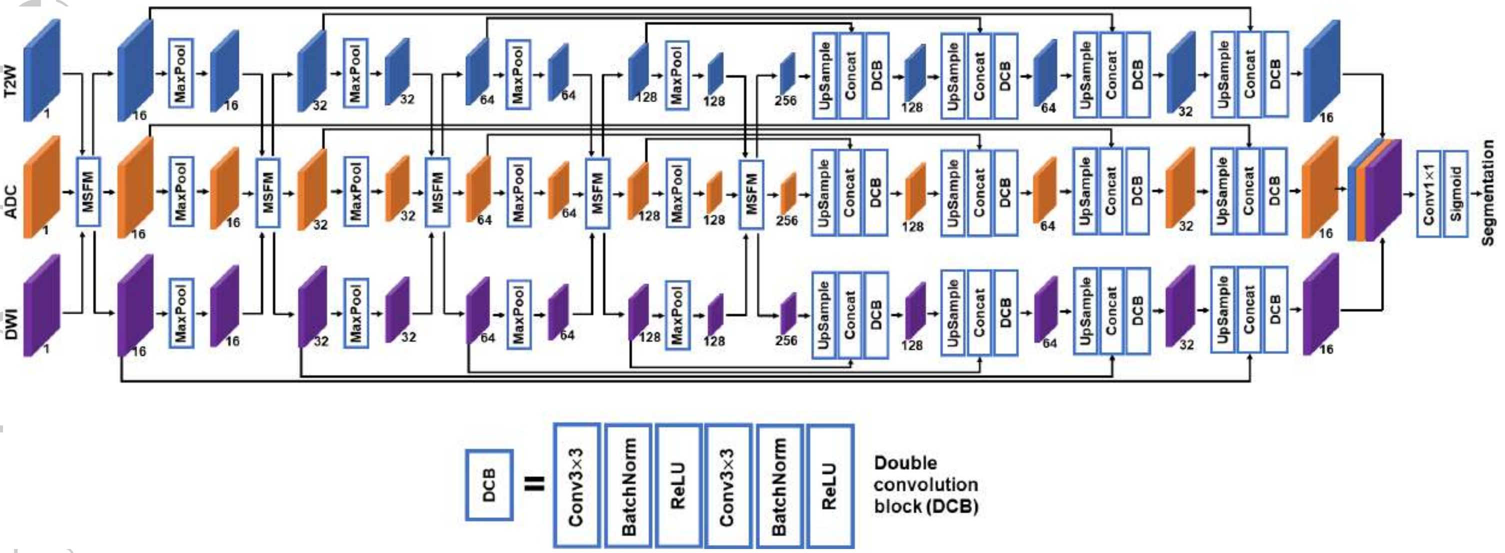
4. Architecture comparison of (a) single-stream and (b) multistream networks with and without the multistream encoder module (represented by the red box)

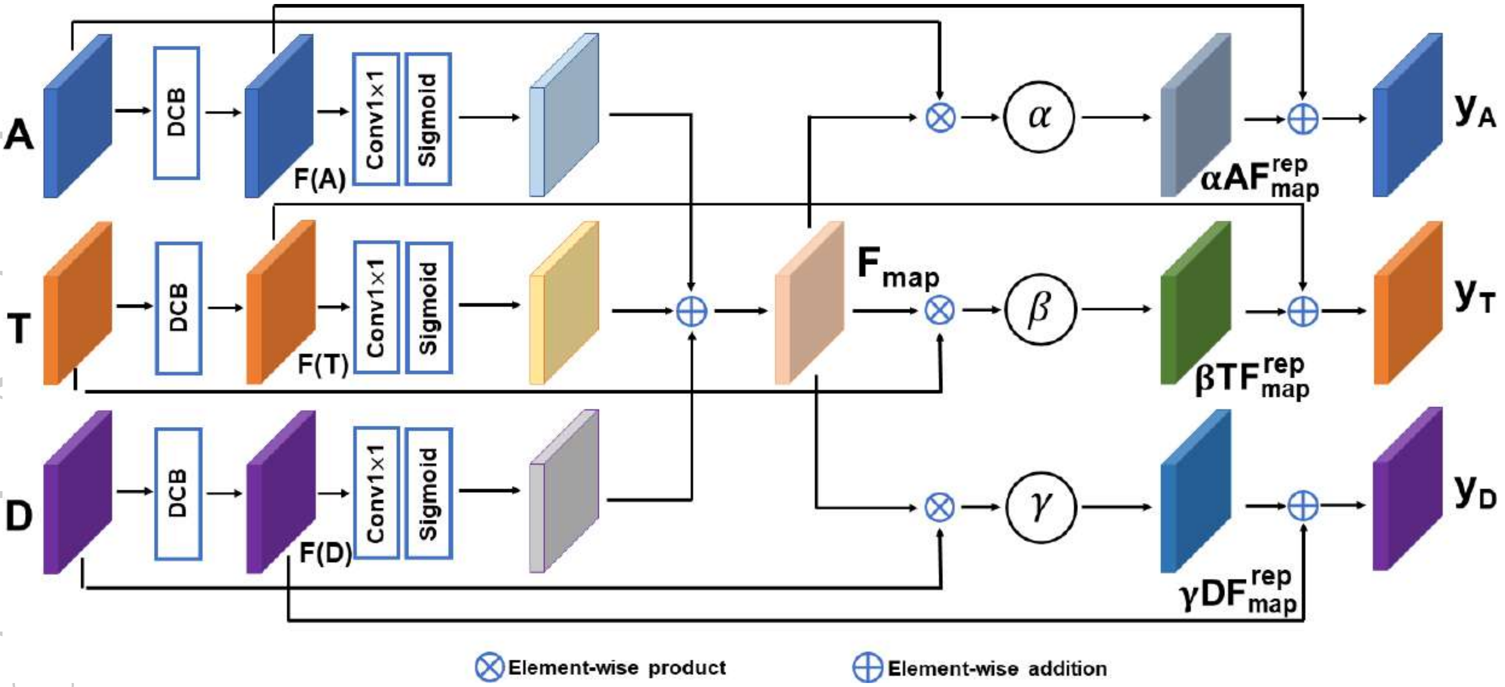
5. Comparison of segmentation results generated without (second row) and with (third row) the patch-based loss function $\mathcal{L}_{patch}$ for four example prostate images. Orange and green arrows point to false positive and false negative segmentations for $\mathcal{L}_{global}$ setting, which were corrected with the introduction of $\mathcal{L}_{patch}$.

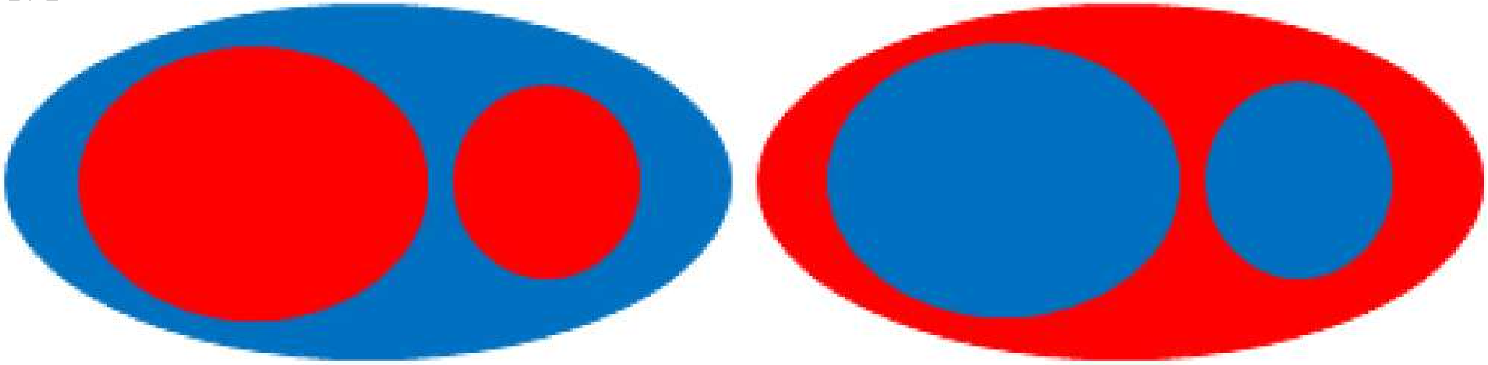
6. Comparison of feature maps obtained at different levels (a) with and (b) without the multistream fusion encoder. Specifically, feature maps in (a) were generated with the MSFusion+Tuning+$\mathcal{L}_{patch}$ setting, whereas those in (b) were generated with the MS setting, described in Sec. III.B.

7. Performance of several CNN-based segmentation methods on four example images. Each row shows the segmentation results obtained using different methods. Segmented lesions are highlighted in red. Orange arrows point to false positive segmentation. Green arrows point to manually segmented lesions delineated by the proposed method, but not segmented by most methods in comparison.

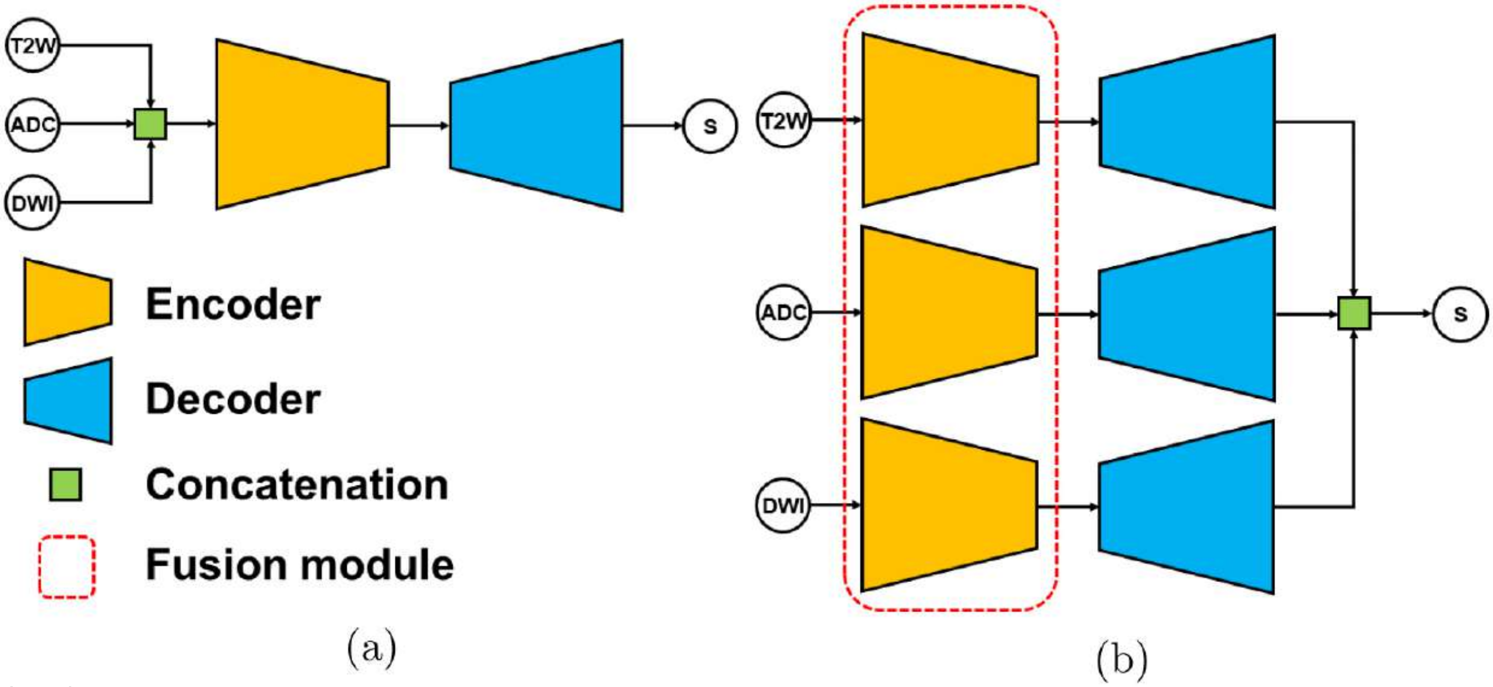
8. Sensitivity of the proposed MSFusion-UNet and the manual segmentation performed by the second radiologist as functions of margin.

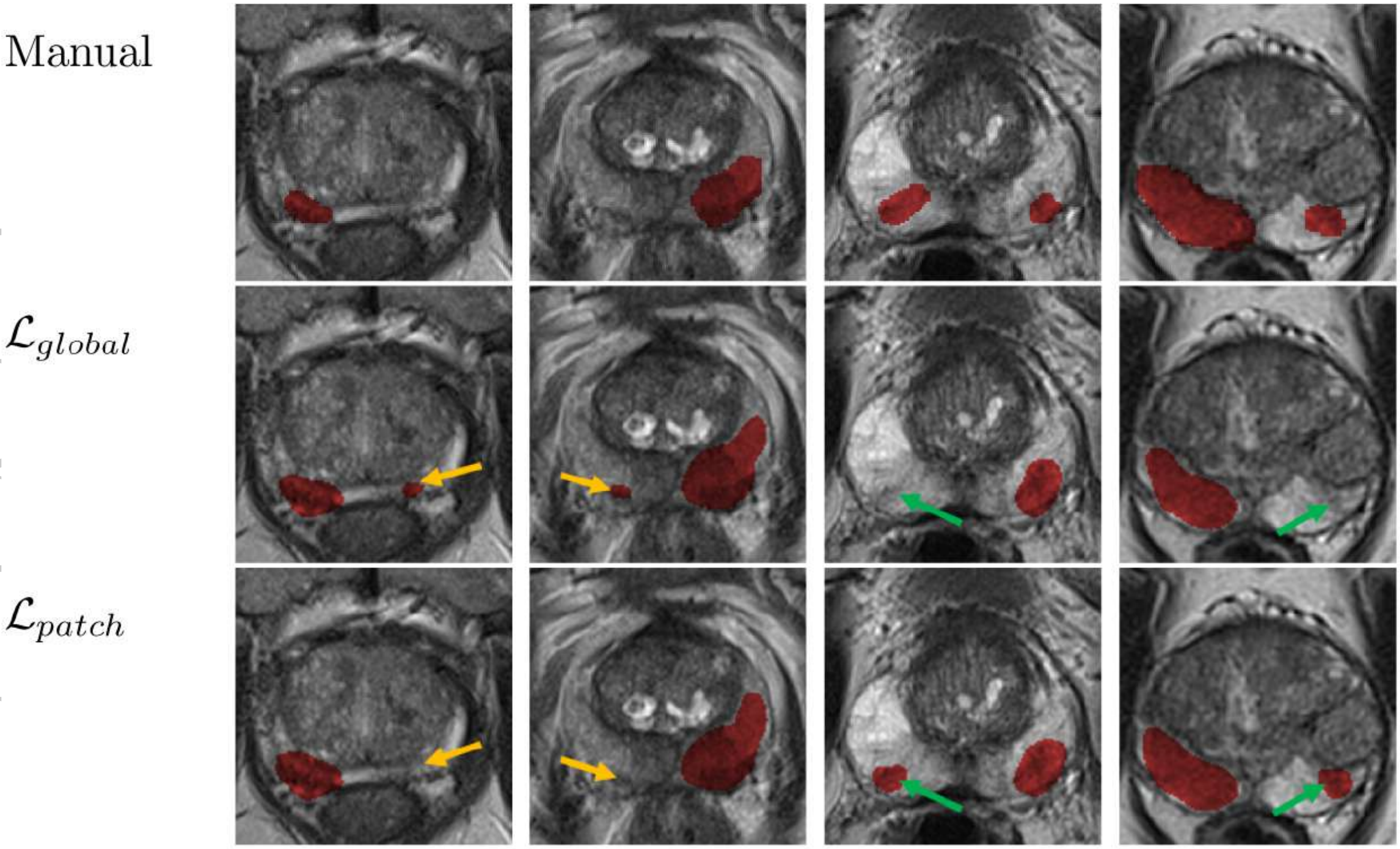

T2W
ADC
DWI
1
MSFM
16
MaxPool
32
64
128
256
UpSample
Concat
DCB
Conv1×1
Sigmoid
Segmentation
DCB
=
Conv3×3
BatchNorm
ReLU
Conv3×3
BatchNorm
ReLU
Double convolution block (DCB)

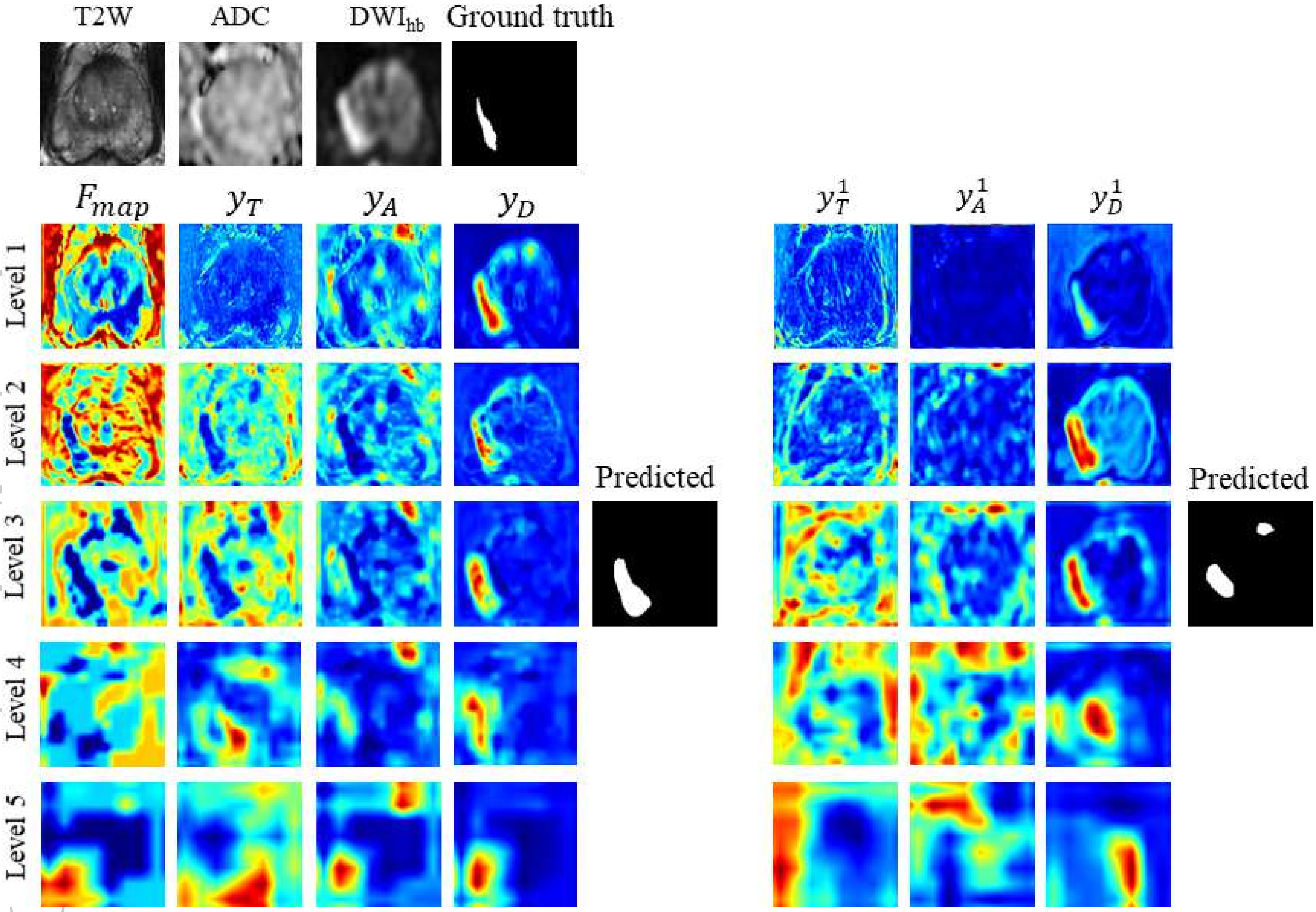

A
DCB
F(A)
Conv1×1
Sigmoid
α
$\alpha A F_{map}^{rep}$
y_A
T
DCB
F(T)
Conv1×1
Sigmoid
F_{map}
β
$\beta T F_{map}^{rep}$
y_T
D
DCB
F(D)
Conv1×1
Sigmoid
γ
$\gamma D F_{map}^{rep}$
y_D
Element-wise product
Element-wise addition

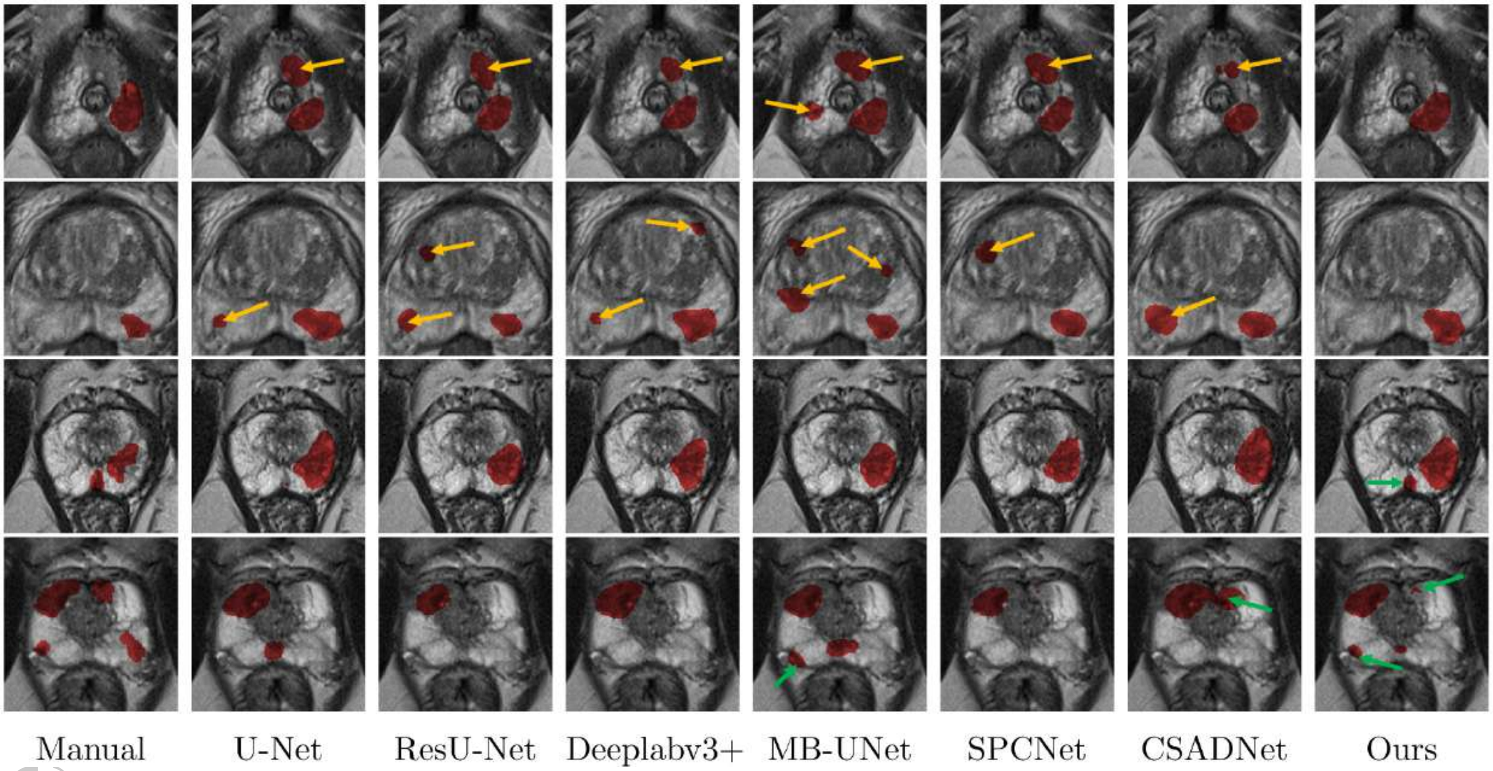


(a) Recall $= 0.5$ (b) Precision $= 0.5$


T2W
ADC
DWI
S
Encoder
Decoder
Concatenation
Fusion module
(a)
T2W
ADC
DWI
S
(b)


Manual
$\mathcal{L}_{global}$
$\mathcal{L}_{patch}$


T2W
ADC
DWI_{hb}
Ground truth
F_{map}
y_T
y_A
y_D
y_T^1
y_A^1
y_D^1
Level 1
Level 2
Level 3
Level 4
Level 5
Predicted
Predicted


Manual
U-Net
ResU-Net
Deeplabv3+
MB-UNet
SPCNet
CSADNet
Ours

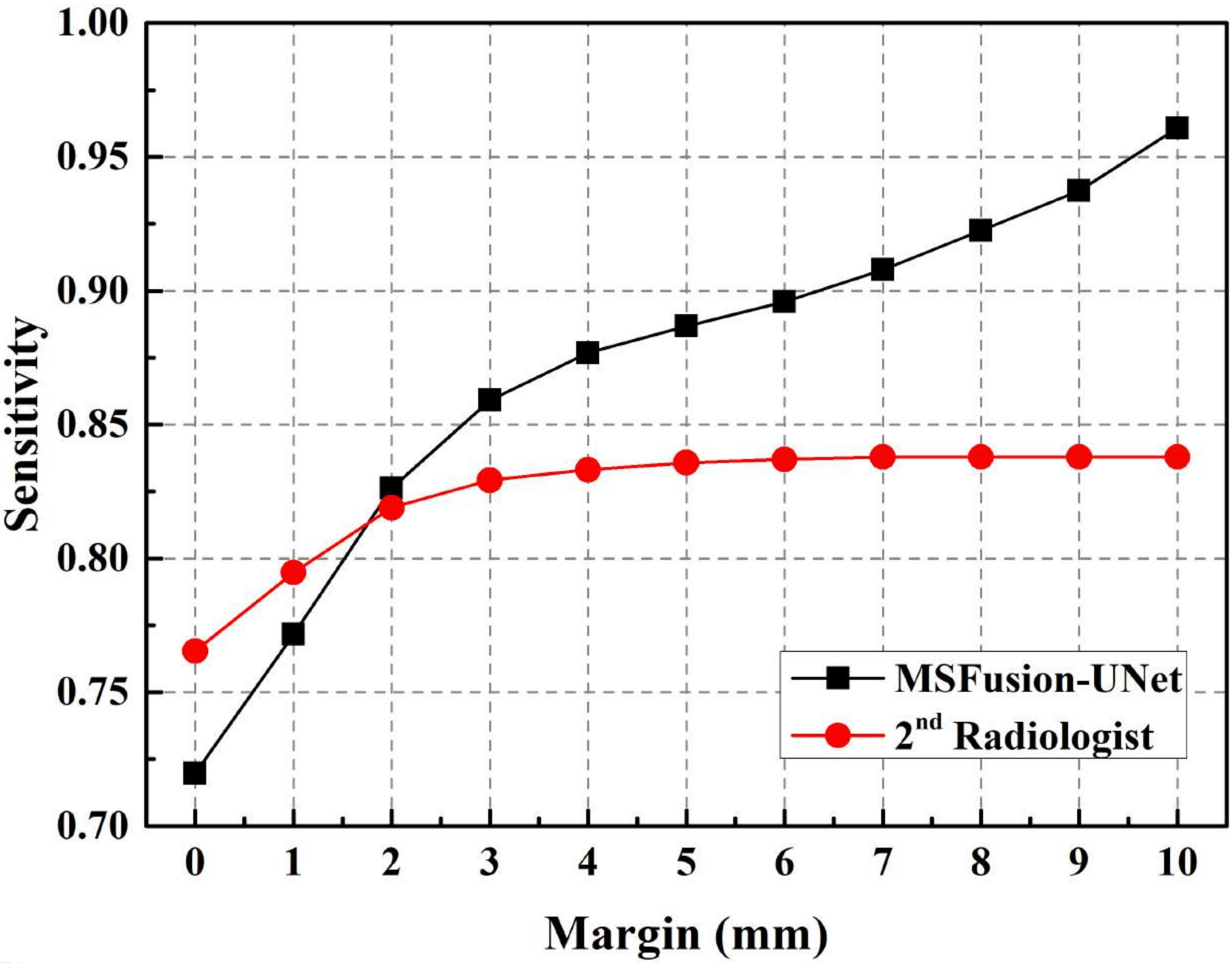

1.00
0.95
0.90
0.85
0.80
0.75
0.70
Sensitivity
0
1
2
3
4
5
6
7
8
9
10
Margin (mm)
MSFusion-UNet
2nd Radiologist

Table 1: MRI scanning parameters for the datasets from Prince of Wales Hospital (PWH) and University College London (UCL) Hospital

	T2W	DWI
	PWH Dataset	
TR	3.3s	2.4s
TE	90ms	84ms
Slice thickness	3mm	3mm
In-plane dimensions	0.45×0.45mm	1.25×1.25mm
High b-value	-	1600
	UCL Datasets	
TR	5.2s	2.2s
TE	92ms	98ms
Slice thickness	0.82-1mm	3-5mm
In-plane dimensions	0.45×0.45mm to 1.31×1.31mm	0.75×0.75mm to 3.41×3.41mm
High b-value	-	1400, 2000

Table 2: Testing performance of the proposed method for different c settings

c	Area-based		Lesion-level			Patient-level		
	DSC	SEN	Precision	Recall	F_1	Precision	Recall	F_1
0	54.5±21.5	69.1±24.4	73.2	91.9	81.5	81.0	94.7	87.3
0.25	52.8±22.0	66.8±25.8	74.6	93.2	82.8	80.2	95.9	87.4
0.5	54.0±21.2	70.5±24.5	73.6	93.2	82.2	80.6	95.9	87.6
0.75	52.9±21.8	67.0±25.5	73.8	91.3	81.6	79.8	94.0	86.3
1	53.6±21.1	69.6±24.0	73.0	91.9	81.4	78.8	95.1	86.2

Table 3: Testing performance of the proposed method for different patch size settings

S_p	Area-based		Lesion-level			Patient-level		
	DSC	SEN	Precision	Recall	F_1	Precision	Recall	F_1
16	54.0±21.3	67.5±24.6	72.4	90.7	80.5	78.5	93.3	85.3
32	54.0±21.2	70.5±24.5	73.6	93.2	82.2	80.6	95.9	87.6
64	53.7±21.8	70.1±24.3	73.0	93.8	82.1	79.2	96.4	87.0

Table 4: Ablation study on the individual contributions of the multistream fusion encoder module (MSFusion), the adaptive weights for different image modalities (Tuning) and the patch-based loss function (L_{patch}). The DSC, sensitivity, recall and precision are reported in percentage.

Model	Area-based		Lesion-level			Patient-level		
	DSC	SEN	Precision	Recall	F_1	Precision	Recall	F_1
Baseline	51.8±21.6	59.7±25.6	71.4	85.7	77.9	77.3	89.9	83.1
MS	52.1±22.7	60.0±26.1	72.3	90.1	80.2	77.7	93.6	84.9
MSFusion	54.0±22.5	64.7±26.4	73.2	90.7	81.0	80.2	94.3	86.7
MSFusion+Tuning	54.5±21.5	69.1±24.4	73.2	91.9	81.5	81.0	94.7	87.3
MSFusion+Tuning+L_{patch}	54.0±21.2	70.5±24.5	73.6	93.2	82.2	80.6	95.9	87.6

Table 5: Performance of the proposed and previous segmentation models quantified by area-based, lesion-level and patient-level metrics on UCL dataset. The DSC, sensitivity, recall and precision are reported in percentage.

Method	Area-based		Lesion-level			Patient-level		
	DSC	SEN	Precision	Recall	F_1	Precision	Recall	F_1
U-Net[38]	51.8±21.6	59.7±25.6	71.4	85.7	77.9	77.3	89.9	83.1
ResU-Net[47]	51.3±22.4	61.3±26.0	68.9	90.1	78.1	75.9	93.1	83.6
Deeplabv3+[48]	52.1±23.1	60.0±27.1	68.1	85.7	75.9	76.8	88.5	82.3
MB-UNet[25]	51.4±21.2	62.8±25.0	73.1	90.1	80.7	79.5	93.9	86.1
SPCNet[26]	50.7±23.5	56.8±26.9	62.3	88.8	73.2	72.4	92.8	81.3
CSADNet[28]	50.1±23.7	59.6±28.7	71.7	87.0	78.6	76.0	90.2	82.5
Ours (MSFusion-UNet)	54.0±21.2	70.5±24.5	73.6	93.2	82.2	80.6	95.9	87.6

Table 6: Inference time of the different segmentation models expressed in milliseconds.

Method	Inference time (ms)
U-Net	3.8
ResU-Net	6.6
Deeplabv3+	17.3
MB-UNet	13.7
SPCNet	10.4
CSADNet	18.3
Ours (MSFusion-UNet)	11.8

Table 7: Performance of the proposed method on PWH dataset. The DSC, sensitivity, recall and precision are reported in percentage.

Method	Area-based		Lesion-level			Patient-level		
	DSC	SEN	Precision	Recall	F_1	Precision	Recall	F_1
Pre-trained	52.6±22.2	78.6±28.1	51.5	96.7	67.2	58.7	97.9	73.4
Fine-tuned	66.3±27.4	72.0±29.7	78.4	93.3	85.2	82.1	93.8	87.5