Signal, Image and Video Processing (2026) 20:434
https://doi.org/10.1007/s11760-026-05429-9

ORIGINAL PAPER

A multitask deep representation dedicated to localize prostate cancer-related lesions on multimodal bpMRI sequences

Camilo González[1] · Juan A. Olmos[1,3] · Catalina de Valencia[2] · Fabio Martínez[1]

Received: 5 August 2025 / Revised: 1 May 2026 / Accepted: 6 May 2026 / Published online: 17 June 2026
© The Author(s) 2026

Abstract

Prostate cancer is the second most common cancer in men worldwide. Early detection is critical for reducing deaths and improving patients' quality of life. Bi-parametric magnetic resonance imaging (bpMRI) has emerged as a potential alternative for supporting the early detection, diagnosis, and screening of prostate cancer. The inherent ambiguity and shape variability of prostate cancer lesions in bpMRI present a major challenge for automated detection systems. To address this, we propose a multitask deep representation that explicitly leverages lesion boundary information to improve localization. Unlike traditional methods that treat localization or segmentation as independent tasks, our approach integrates a segmentation branch that forces the network to learn fine-grained contour details. This synergy results in a more robust feature representation, improving the model's ability to distinguish clinically significant lesions from surrounding tissue. Our approach was validated on a public dataset of 3343 slices from 1295 bpMRI studies. The proposed model achieved an Average Precision (AP@0.5) of 0.59, a precision of 0.66, and a recall of 0.56. This represents a 22.9% improvement in AP@0.5 compared to a standard localization-only network. Furthermore, compared with PI-CAI reports, our model achieved an AP@0.1 of 0.72 under the commonly used criterion. Additionally, within our evaluation protocol, the model achieved an AP@0.5 of 0.59 as a complementary stricter evaluation. In an additional validation using annotations from an expert radiologist as ground truth, the model achieved a precision of 0.79. Our findings indicate that forcing a network to be boundary-aware through segmentation is an effective strategy for improving lesion localization.

Keywords Prostate cancer · Localization · bpMRI · csPCa lesion · Multimodal

1 Introduction

Prostate cancer (PCa) is the second most common cancer in men, with a total of 1.4 million new cases and 300,000 associated deaths reported in 2020 [1]. In the clinical workflow, the most commonly used methods are the prostate-specific antigen (PSA) test and the digital rectal exam (DRE) [2]. However, the PSA test has demonstrated a low specificity of close to 20%, and the DRE evaluation may miss tumors in different areas of the prostate gland, thus overlooking clinically significant PCa (csPCa) cases [3, 4]. Early detection of PCa remains a critical challenge for designing treatments that improve patient prognosis by reducing the number of aggressive lesions and associated deaths. In this context, Multiparametric Magnetic Resonance Imaging (mp-MRI) is a promising alternative to support early diagnosis before biopsy [5]. These sequences allow for the characterization of lesions, even in regions far from the rectal wall and transitional zone [6]. Nonetheless, acquiring mp-MRI

Camilo González, Juan A. Olmos, Catalina de Valencia and Fabio Martínez have contributed equally to this work.

✉ Fabio Martínez
famarcar@saber.uis.edu.co

Camilo González
camilo2180065@saber.uis.edu.co

Juan A. Olmos
jaolmosr@correo.uis.edu.co

Catalina de Valencia
catalina.devalencia@gmail.com

[1] Biomedical Imaging, Vision and Learning Laboratory (BIVL2ab), Universidad Industrial de Santander (UIS), Bucaramanga 680001, Colombia

[2] Fundación Oftalmológica de Santander (FOSCAL), Floridablanca 681004, Colombia

[3] ENSTA, Institut Polytechnique de Paris, Palaiseau 91120, France

is time-consuming, and the inclusion of contrast-enhanced (DCE) imaging often does not provide a significant diagnostic gain [7]. Consequently, Bi-parametric MRI (bpMRI), without DCE, has emerged as a faster alternative with comparable diagnostic accuracy [8]. However, the benefit of MRI is often limited by the subjectivity of expert interpretation, which can lead to inter-reader variability and diagnostic errors [9, 10]. These challenges, compounded by logistical issues such as staff shortages, have spurred the development of computational tools to provide support and strengthen the screening process [11, 12]. Early approaches based on engineered features [13] showed promise but were often sensitive to variations across scanners and prostate zones. The field subsequently shifted towards end-to-end deep learning models for localization and segmentation tasks [14–16], yet these strategies still struggle to capture the shape variability of lesions where precise boundary delineation is critical. This work introduces a multitask framework that improves csPCa lesion localization by making the network explicitly boundary-aware. We hypothesize and demonstrate that by integrating a segmentation task not as an independent objective, but as a synergistic guide, the network learns a more robust feature representation. This forces the model to focus on fine-grained contour details, enhancing its ability to localize clinically significant lesions in bpMRI sequences.

2 Related works

The approach to csPCa lesion localization has evolved from early methods using engineered features with classifiers [13] to end-to-end deep learning networks. Initial deep models often used only T2-weighted images [17], while later multitask frameworks incorporated auxiliary tasks such as prostate segmentation or image alignment [14]. The release of the PI-CAI dataset spurred the development of more advanced methods, including CNNs like Z-SSMNet [15], zonal-aware U-Nets [18], and transformer-based architectures [19]. Recent studies have also explored residual attention and bioimage sensor-driven attention strategies for prostate MRI [20, 21], as well as SAM-based segmentation and federated privacy-preserving learning in related medical-image settings [22, 23]. However, most PI-CAI-related methods primarily rely on segmentation, which often requires complex processing. Our work, in contrast, employs a detector-centric framework based on the YOLO architecture, which is well-established for object detection [24]. Recent advancements have introduced Non-Maximum Suppression (NMS)-free, one-to-one assignment schemes to simplify instance segmentation [25]. While effective for general objects, these methods may struggle with the ill-defined boundaries of prostate lesions in bpMRI. To deal with prostate variability and poor definition of organ and lesion boundaries, it is demanding that detector approaches also include segmentation steps, that may focus interest at each pixel of regions, considering remarkable changes in appearance and morphology. Also, detectors may consider to weight boundaries of regions of interest as an internal guide to improve localization.

3 Proposed approach

Our method adapts YOLOv5 by incorporating a synergistic segmentation branch. YOLOv5 was selected as a stable detector-centric baseline with SPPF/PANet multi-scale aggregation, suitable for evaluating the proposed multitask localization strategy. For comparative analysis, we use two configurations of this architecture: (*Loc+Seg*) and a localization-only baseline (*Loc*).

3.1 Localization network

The localization network processes a concatenated multimodal bpMRI slice, $\mathbf{x} \in \mathbb{R}^{3\times H\times W}$ (T2W, ADC, DWI), through three main stages: an encoder, a neck, and a prediction head. First, the encoder ($\mathbf{\Phi}$), a standard convolutional network, extracts a bank of deep features from the input slice: $\mathbf{F}^k = f_k(\mathbf{x} * \mathbf{\Phi}_k)$, where f_k is a nonlinear activation function.

The final feature maps from the encoder, $\mathbf{F}_L$, are processed by a Spatial Pyramid Pooling Fusion (SPPF) module to create a compact and robust multiscale representation. Next, the network's neck, composed of a decoder ($\mathbf{\Phi}'$) and a Path Aggregation Network (PANet) [26], fuses the multiscale features. The decoder upsamples the feature maps while integrating intermediate features from the encoder, producing a set of representations $\{\mathbf{G}_i\}$. The PANet then adds an extra bottom-up aggregation path to further enhance the feature fusion, improving the model's ability to detect lesions of varying sizes. Finally, the prediction head generates anchor-based predictions from the fused feature maps. For each potential detection, it outputs a vector $\mathbf{Y}_{i,j,a} = [\delta_x, \delta_y, \delta_w, \delta_h, \rho_{\text{obj}}, \{\rho^c_{\text{cls}}\}_{c=1}^{C}]$, containing the bounding box offsets, an objectness score, and a lesion confidence score. The raw set of predicted boxes, $\mathcal{B}$, is then refined using Non-Maximum Suppression (NMS) to produce the final detections, $\mathcal{B}_{\text{sel}}$. The IoU metric is computed as $\text{IoU}(\mathbf{B}_i, \mathbf{B}_j) = \frac{|\mathbf{B}_i \cap \mathbf{B}_j|}{|\mathbf{B}_i \cup \mathbf{B}_j|} \in [0, 1]$.

3.2 Segmentation branch

To enhance localization, we use a YOLACT-inspired prototype-coefficient branch [27] as auxiliary boundary-aware supervision rather than as the final clinical segmentation output.

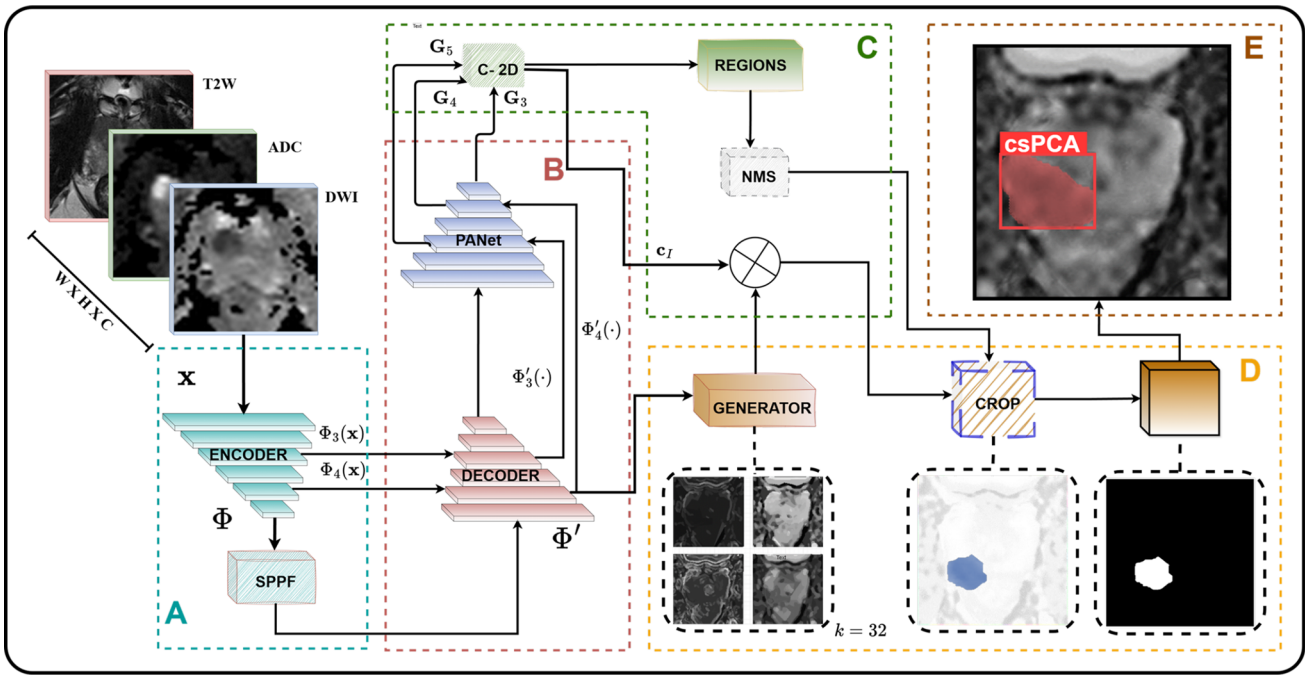


Fig. 1 Structure of the implemented network for csPCa lesion localization task: **(A)** Feature extraction. **(B)** Multiscale grouping representation. **(C)** Mechanism for region prediction (Head). **(D)** Segmentation network. **(E)** Example of final csPCa region prediction

Operating in parallel with the localization head, it consists of two components. First, a Prototype Generation network uses a Fully Convolutional Network to produce a set of $k = 32$ prototype masks, denoted as $\mathcal{P} = \{\mathbf{P}_j\}_{j=1}^{k}$. Second, a Mask Coefficient Prediction head predicts k mask coefficients organized in a vector $\mathbf{c}_I \in \mathbb{R}^k$, for each instance I that survives NMS. The final mask for the instance, $\mathbf{M}_I$, is generated by a linear combination of prototypes and their corresponding coefficients: $\mathbf{M}_I = \sum_{j=1}^{k} c_{I,j}\mathbf{P}_j$. This mask is then processed with a sigmoid function and binarized. The entire branch is trained jointly with the localization network, forcing the shared feature extractor to learn boundary-aware features (Figure 1(D)).

3.3 Learning scheme

The network is trained end-to-end by optimizing $\mathcal{L}_{\text{total}}$. As this is a single-class detection problem, the total loss is a weighted sum of three components:

$$\mathcal{L}_{\text{total}} = \gamma_{\text{box}}\mathcal{L}_{\text{box}} + \gamma_{\text{obj}}\mathcal{L}_{\text{obj}} + \gamma_{\text{seg}}\mathcal{L}_{\text{seg}} \quad (1)$$

where γ_{box} and γ_{obj} are hyperparameters that balance the localization and objectness tasks. The box loss ($\mathcal{L}_{\text{box}}$) is based on the Complete Intersection over Union (CIoU) metric. The objectness loss ($\mathcal{L}_{\text{obj}}$) is a weighted binary cross-entropy (BCE) loss. Finally, the segmentation loss ($\mathcal{L}_{\text{seg}}$) is a standard BCE loss.

4 Experimental setup

4.1 Dataset

The dataset is derived from the public *PI-CAI (Prostate Imaging: Cancer AI)* consortium, a multicentric cohort of 1,500 cases with histopathology-verified annotations for clinically significant prostate cancer (csPCa) [12]. Each study, composed of T2-weighted (T2W), Diffusion-weighted (DWI), and Apparent Diffusion Coefficient (ADC) images, underwent a standardized pre-processing pipeline. This included resampling to a voxel spacing of $0.5 \times 0.5 \times 3.0\ mm$, co-registration to the T2W sequence, and cropping a 150x150 pixel region of interest around the prostate gland centroid. For model training, these 2D slices were normalized to a $[0 - 1]$ range and their channels were combined. To create the training dataset, we curated 1,282 slices containing csPCa annotations and 2,061 benign slices selected from negative cases. Ground truth labels were then generated for the csPCa cases. To optimize training by capturing a more comprehensive imaging context surrounding lesions, the labels for localization and segmentation were standardized through a 30% expansion of the bounding box and a 5-pixel dilation of the mask, respectively. These values were identified as optimal through an ablation study, as established in the results presented in Table 2.

Table 1 Comparison of different bpMRI input configurations using the Localization network **without (top)** and **with (bottom)** segmentation branch

Input	Precision↑	Recall↑	AP@0.5↑	AP@0.5-.95↑
Localization				
T2W	0.39 ± 0.06	0.24 ± 0.04	0.21 ± 0.04	0.07 ± 0.02
ADC	0.54 ± 0.06	0.42 ± 0.06	0.42 ± 0.06	0.16 ± 0.03
DWI	0.49 ± 0.05	0.41 ± 0.04	0.36 ± 0.05	0.13 ± 0.02
ADC-T2W-ADC	0.60 ± 0.08	0.41 ± 0.05	0.44 ± 0.06	0.18 ± 0.03
ADC-DWI-ADC	0.56 ± 0.07	0.49 ± 0.06	0.47 ± 0.05	0.18 ± 0.03
T2W-ADC-DWI	$\mathbf{0.59 \pm 0.05}$	$\mathbf{0.50 \pm 0.07}$	$\mathbf{0.48 \pm 0.06}$	$\mathbf{0.18 \pm 0.03}$
Localization & Segmentation				
T2W	0.33 ± 0.04	0.23 ± 0.05	0.18 ± 0.03	0.06 ± 0.01
ADC	0.59 ± 0.07	0.50 ± 0.03	0.49 ± 0.06	0.21 ± 0.03
DWI	0.57 ± 0.05	0.45 ± 0.06	0.45 ± 0.05	0.17 ± 0.03
ADC-T2W-ADC	0.62 ± 0.05	0.57 ± 0.02	0.56 ± 0.03	0.26 ± 0.03
ADC-DWI-ADC	0.65 ± 0.02	0.56 ± 0.05	0.58 ± 0.04	0.25 ± 0.03
T2W-ADC-DWI	$\mathbf{0.66 \pm 0.05}$	$\mathbf{0.56 \pm 0.06}$	$\mathbf{0.59 \pm 0.05}$	$\mathbf{0.26 \pm 0.03}$

4.2 Network Configuration

For comparative analysis, two configurations of the proposed architecture were used: our full multitask model (*Loc+Seg*) and a localization-only baseline (*Loc*). Both models were initialized with weights pre-trained on the Microsoft COCO dataset [28]. The models were trained for 100 epochs on a single NVIDIA GeForce RTX 4090 GPU using the Adam optimizer. Key training hyperparameters included a learning rate of 0.01, momentum of 0.937, and a weight decay of 0.0005. The loss function weights for the detection tasks were set to $\gamma_{box} = 0.05$, $\gamma_{obj} = 1.0$ and $\gamma_{seg} = 1.0$, following the framework's default values. Data augmentation techniques included mosaic, translation, scaling, and horizontal flipping.

5 Evaluation and results

Experiments were conducted using the publicly available PI-CAI challenge training cohort [12]. Model training and evaluation followed a stratified 5-fold cross-validation protocol, with folds defined at the patient/study level to avoid data leakage; all metrics were computed at the slice level without 3D aggregation. Performance was assessed using standard metrics, including precision, recall, Average Precision at an IoU threshold of 0.5 (AP@0.5), and the stricter AP@0.5-0.95 metric. The results correspond to the mean and standard deviation of these metrics across the five folds.

First, we analyzed the performance of the baseline localization-only network using different input sequence combinations, with detailed results in Table 1. Key findings indicate that while ADC was the strongest single modality (precision 0.54 ± 0.06), multimodal inputs consistently performed better. The ADC-T2W-ADC combination achieved the highest precision (0.60 ± 0.08), while the full T2W-ADC-DWI input yielded the best overall balance with the highest AP@0.5 (0.48 ± 0.06). However, under the stringent AP@0.5-0.95 metric, the performance of all multimodal configurations converged to an identical score (0.18 ± 0.03), suggesting a performance ceiling for this baseline approach. The primary contribution of our work is demonstrated in the second part of Table 1, which details the performance of our full *Loc+Seg* model. The addition of the synergistic segmentation branch improved performance across all configurations. The most significant benefit was observed in the complete bpMRI model (T2W-ADC-DWI), where the segmentation branch boosted the AP@0.5 score from 0.48 ± 0.06 to 0.59 ± 0.05—a substantial 22.9% improvement. Notably, the AP metrics showed the highest percentage gains (e.g., 44.4% in AP@0.5-0.95), underscoring that the boundary information provided by the segmentation task is particularly effective at improving localization precision. As a complementary analysis, the mask branch achieved a Dice coefficient of 0.66 ± 0.02 against the auxiliary lesion-mask targets on lesion-positive slices using the highest-confidence predicted mask after NMS, supporting its role as spatial regularization rather than as a clinical segmentation output.

To further validate the impact of our synergistic design, we compared the complete bpMRI *Loc+Seg* model against its *Loc* baseline across a range of IoU thresholds (τ from 0.1 to 0.9). The results, illustrated in Figure 2, demonstrate a clear and consistent performance advantage for the *Loc+Seg* model. While both models show an expected performance decline as the localization requirements become stricter (higher τ), the decline for the baseline *Loc* model is significantly steeper, particularly for recall and AP. This is most evident at moderate to high thresholds ($\tau \geq 0.5$), where the *Loc+Seg* model maintains a more stable and superior

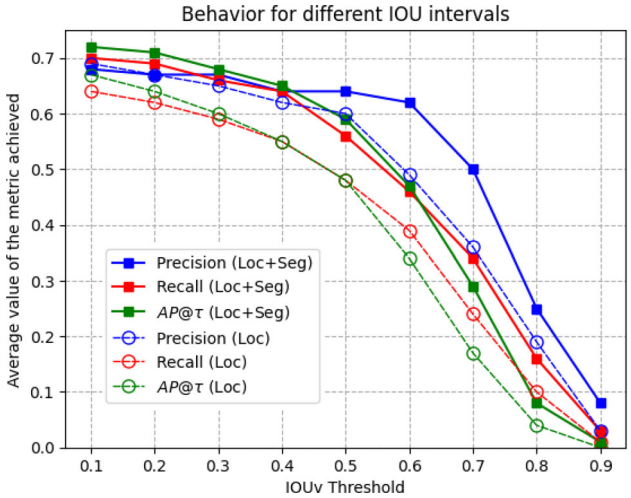


Fig. 2 Performance comparison of the *Loc+Seg* and *Loc* models varying IoU thresholds (τ)

performance. These findings confirm that the boundary-aware features learned via the segmentation branch provide enhanced stability against localization inaccuracies, a crucial factor for reliable performance under stringent clinical evaluation.

5.1 Ablation study

We performed a sensitivity analysis of the *Loc+Seg* model with respect to different settings in the bounding-box expansion and the dilation of the mask ground truths (see Table 2). First, we varied the bounding box expansion using 0%, 15%, and 30%, observing that 30% was the optimal setting. Subsequently, we evaluated mask dilations of 0, 2, 5, and 10 pixels. The results showed that a 5-pixel dilation yielded the best performance across all metrics. In this sense, we observe that incorporating additional imaging contextual information from the areas surrounding the lesions during training enhances the performance of the proposed network.

In addition, we performed a sensitivity analysis on the loss function weights. We compared the default configuration and varied the values of the box weight (γ_{box}), objectness weight (γ_{obj}), and segmentation weight (γ_{seg}). As shown in Table 3, the model's performance remains stable with variations in the segmentation weight (γ_{seg}), achieving results comparable to the default configuration. Conversely, the model's performance significantly deteriorates when the box weight (γ_{box}) is modified. Furthermore, a slight increase in the objectness weight (γ_{obj}) results in a marginal improvement, even surpassing the default configuration. However, a slight decrease in this weight leads to a marginal decline in the model's performance, indicating its sensitivity. In this sense, the segmentation weight demonstrated greater robustness.

5.2 Visualization of learned representations

To assess differences in learned representations, we visualized individual channel activations from the baseline (*Loc*) and proposed multitask (*Loc+Seg*) models. As shown in Figure 3, we extracted feature maps from three stages: *early* backbone stage (layer 1), *mid* neck top-down path (layer 13), and *final* neck bottom-up path (layer 23). The visualizations reveal progressive differences across network depth. At the early stage, both models capture anatomical edges, but *Loc+Seg* shows different activation patterns in intensity and spatial distribution. At the mid stage, the divergence becomes more pronounced: *Loc+Seg* produces focused activations on the prostate region, while *Loc* shows diffuse responses with weak representation of the lesion area. At the final stage, *Loc* exhibits scattered activation beyond lesion boundaries, whereas *Loc+Seg* generates brighter, more compact activations aligned with the lesion edges. Although in this particular case both predictions are close to each other (see Panel C), the learned representations suggest that the segmentation task guides the network from early stages toward a progressively more boundary-aware representation, explaining a better localization performance.

Table 2 Ablation of ground-truth. The analysis involves a comparison of performance across configurations, including the expansion of bounding boxes and the dilation of mask delineations. The optimal parameters are highlighted in bold

Parameter variation	Precision↑	Recall↑	AP@0.5↑	AP@0.5-.95↑
Bounding Box Expansion (%)				
0% (No Expansion)	0.48 ± 0.04	0.35 ± 0.06	0.31 ± 0.05	0.07 ± 0.01
15%	0.58 ± 0.08	0.46 ± 0.06	0.45 ± 0.07	0.15 ± 0.03
30% (Reported)	**0.59 ± 0.05**	0.50 ± 0.07	0.48 ± 0.06	0.18 ± 0.03
Segmentation Mask Expansion (pixels)				
0 (No Expansion)	0.48 ± 0.06	0.33 ± 0.06	0.28 ± 0.06	0.06 ± 0.02
2	0.58 ± 0.03	0.52 ± 0.01	0.50 ± 0.02	0.17 ± 0.02
5 (Reported)	**0.66 ± 0.05**	**0.56 ± 0.06**	**0.59 ± 0.05**	**0.26 ± 0.03**
10	0.54 ± 0.06	0.45 ± 0.05	0.41 ± 0.07	0.10 ± 0.02

Table 3 Ablation Study on loss function weights. The table presents the default configuration used in the experiments and the corresponding results under variations of each loss weight

Configuration	Precision↑	Recall↑	AP@0.5↑	AP@0.5-.95↑
Default Configuration				
$\gamma_{box} = 0.05, \gamma_{obj} = 1.0, \gamma_{seg} = 1.0$	**0.66 ± 0.05**	**0.56 ± 0.06**	**0.59 ± 0.05**	**0.26 ± 0.03**
Box Weight Variation				
$\gamma_{box} = 0.1, \gamma_{obj} = 1.0, \gamma_{seg} = 1.0$	0.38 ± 0.04	0.28 ± 0.04	0.23 ± 0.03	0.05 ± 0.02
Objectness Weight Variation				
$\gamma_{obj} = 0.5, \gamma_{box} = 0.05, \gamma_{seg} = 1.0$	0.62 ± 0.05	0.54 ± 0.06	0.55 ± 0.05	0.24 ± 0.03
$\gamma_{obj} = 2.0, \gamma_{box} = 0.05, \gamma_{seg} = 1.0$	0.67 ± 0.05	0.58 ± 0.06	0.60 ± 0.05	0.27 ± 0.04
Segmentation Weight Variation				
$\gamma_{seg} = 0.5, \gamma_{box} = 0.05, \gamma_{obj} = 1.0$	0.63 ± 0.05	0.57 ± 0.07	0.57 ± 0.06	0.24 ± 0.04
$\gamma_{seg} = 2.0, \gamma_{box} = 0.05, \gamma_{obj} = 1.0$	0.64 ± 0.04	0.58 ± 0.04	0.58 ± 0.05	0.25 ± 0.03

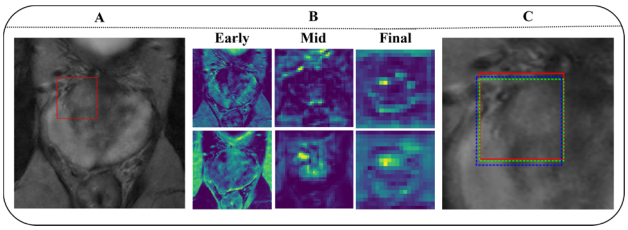


Fig. 3 **Learned representations.** **(A)** Input bp-MRI slice with ground truth annotation. **(B)** Activation maps at early, mid, and final stages for *Loc* (top row) and *Loc+Seg* (bottom row). **(C)** Zoomed comparison of predictions: *Loc+Seg* (green) vs. *Loc* (blue)

Table 4 Performance comparison with state-of-the-art methods using the commonly reported AP@0.1 criterion. AP@0.5 is reported only for the proposed method as a complementary stricter metric

Method	τ=0.1	τ=0.5
U-Net [15]	0.463	-
nnU-Net [15]	0.457	-
Z-SSMNet [15]	0.709	-
Karagoz et al. [18]	0.732	-
ITUNet [19]	0.649	-
PIMed [29]	0.681	-
Proposed *(Loc+Seg)*	**0.72**	**0.59**

5.3 Comparison with the State of the Art

To contextualize our performance, we compared our results with PI-CAI-related reports, as summarized in Table 4. Under the commonly reported AP@0.1 criterion, the proposed *Loc+Seg* model achieved 0.72, which is in the range of previously reported PI-CAI-related methods. Additionally, our model yields an AP@0.5 of 0.59 within our evaluation protocol, which is presented as a complementary stricter evaluation rather than as a direct state-of-the-art ranking.

5.4 Comparison with radiology report

To further assess the model's clinical relevance, we conducted an additional validation on a subset of 73 csPCa cases. For this experiment, manual delineations from an expert radiologist (10+ years of experience) following the PI-RADS v2.1 protocol [30] were used as ground truth. This provided a clinically-oriented benchmark, distinct from the histopathology-based ISUP annotations of the original dataset. When evaluated against this ground truth, our *Loc+Seg* model achieved a precision of 0.79 ± 0.10 and an AP@0.5 of 0.63 ± 0.06. These results indicate a strong correspondence between the model's predictions and the expert's assessment, suggesting its potential as a supportive tool in clinical workflows.

A qualitative analysis of the model's predictions is presented in Figure 4. The examples visually confirm the quantitative findings. Cases A-C illustrate the superior alignment of the *Loc+Seg* model compared to the baseline, with its predictions more closely matching the lesion's true shape. The figure also presents the model's current limitations. A closer look at these failure modes reveals that occasional false positives (Case D) tend to occur in anatomically complex regions like the transitional zone, where benign conditions such as chronic prostatitis or benign prostatic hyperplasia (BPH) can mimic csPCa. Conversely, missed detections (Case E) are typically associated with lesions that present subtle imaging features, such as smaller tumors or lower-grade lesions (e.g., ISUP Grade Group 2). Finally, cases with discrepancies between ISUP and radiologist annotations (Cases F-G) highlight the model's ability to localize the clinically relevant lesion identified by both ground truths, underscoring its potential to handle inter-observer variability.

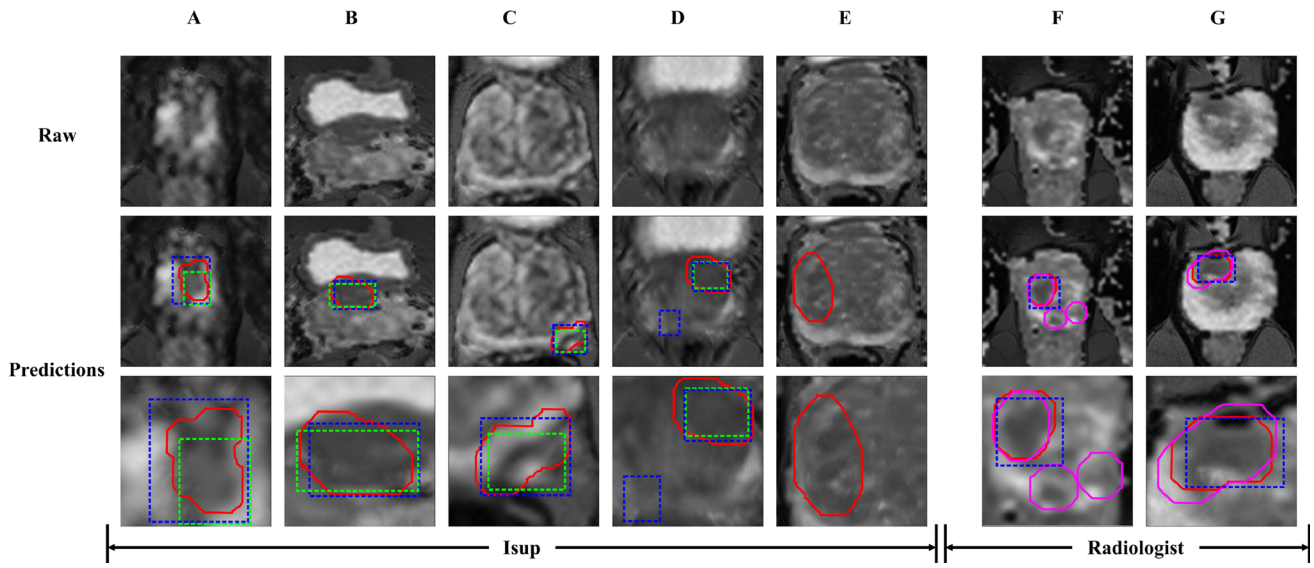


Fig. 4 Qualitative comparison of the baseline *Loc* model (green boxes) and our proposed *Loc+Seg* model (blue boxes). Predictions are compared against two ground truths: histopathology-based ISUP delineations (red) and annotations from an expert radiologist (magenta). The examples show cases of superior alignment (A-C), false positives (D), missed detections (E), and robust performance despite inter-observer variability (F-G)

6 Discussion

This study introduced a multitask localization and segmentation network, demonstrating that a boundary-aware design outperforms a standard localization-only approach. The proposed *Loc+Seg* model surpassed its baseline counterpart *Loc* by 22.9% in AP@0.5. This primary finding supports our central premise that a synergistic segmentation task creates a more robust feature representation, reducing localization ambiguity. This was further evidenced by the model's ability to maintain higher precision under increasingly stringent IoU thresholds, confirming the value of explicit boundary awareness. Our methodology was also consistent with clinical best practices by utilizing complete bpMRI sequences, where the T2W modality provides essential anatomical context, a choice aligned with established PI-RADS v2.1 guidelines [30]. A key contribution of this work was the model's validation against annotations from an expert radiologist following the PI-RADS protocol. Our model achieved a precision of 0.79 on these clinical delineations, suggesting its capacity for independent assessment beyond simply replicating histopathological findings. This aligns with the intended role of such tools as decision support in clinical practice. We also acknowledge the model's limitations, such as missed low-grade lesions (e.g., ISUP 2), which underscores the need for further sensitivity improvements, particularly for the frequent intermediate-risk lesions [31]. Our approach, however, shows promise in bridging discrepancies between ISUP and PI-RADS annotations, enhancing the generalization of csPCa detection.

The analysis of learned representations at different stages of the proposed network provided a qualitative explanation for the observed improvements. From early stages, *Loc+Seg* develops different activation patterns that progressively diverge through the feature pyramid toward boundary-aware filters. This demonstrates that the auxiliary segmentation task reshapes how the network encodes morphological spatial information of lesions, directly translating into superior localization performance. Our approach was also compared with the performance reported for state-of-the-art methods on the PI-CAI challenge hidden validation cohort, as summarized in Table 4. Under this comparison, our model achieved an AP@0.1 of 0.72, placing it close to the top reported performance of 0.73 by Karagoz et al. [18]. It is important to note that these leading methods, such as the U-Net-based approach by Karagoz et al. [18] and the transformer-based ITUNet [19], correspond to sophisticated segmentation networks. In order to compare with standard baseline segmentation models, we included the performance of the U-Net and nnU-Net models [15]. Compared with the reported U-Net and nnU-Net baselines under the AP@0.1 criterion, the proposed hybrid *Loc+Seg* model obtains a higher score, suggesting that the multitask design is beneficial for lesion localization. Moreover, the challenge criterion $\tau = 0.1$ is useful for assessing approximate lesion overlap, but it is less informative regarding spatial agreement. Since AP@0.5 is unavailable for most compared methods, the AP@0.5 value of 0.59 should be interpreted as complementary evidence within our protocol under a higher-overlap criterion, not as a direct comparison against those methods.

Despite the promising results, we identify several avenues for future work. To improve false positive reduction, the model could be enhanced by incorporating clinical data and patient history. Furthermore, regarding the input fusion, while our current approach employs channel stacking, recent collaborative frameworks have demonstrated that reserving modality-specific features while seeking common ground can enhance multimodal representation learning [32]. Adopting such fusion strategies could further refine the feature extraction from T2W, ADC, and DWI sequences to improve localization. Another key extension is to integrate zonal information to improve performance in anatomically complex regions like the transitional zone, where radiological interpretation is notoriously challenging [33]. Additionally, because predictions are generated independently per slice, 3D aggregation will be considered to improve volumetric consistency and reduce duplicate detections across adjacent slices. Recent works have used AI-generated pseudo-labels to increase the number of training annotations, demonstrating improvements in the segmentation of csPCa lesions [18]. We believe that the proposed method could further benefit from these semi-supervised learning strategies. Studies in other areas of medical image analysis have shown the effectiveness of these strategies in producing more reliable pseudo-labels, refining delineations, and employing uncertainty-based aggregation to achieve enhanced contours, particularly in cases involving small tumors [34–36]. Finally, our strategy should be further validated on larger, more diverse datasets.

7 Conclusions

This work introduced and validated a detector-centric multitask framework for the localization of clinically significant prostate cancer in bpMRI. Our central finding is that incorporating a synergistic segmentation branch makes the network explicitly boundary-aware, leading to measurable improvements in localization accuracy and robustness. The clinical relevance of this approach was underscored by its correspondence with annotations from an expert radiologist. This suggests that a synergistic, boundary-aware design is a promising strategy to mitigate inter-reader variability and provide reliable support in clinical workflows. Future work will focus on validating this framework on larger, more diverse patient cohorts and exploring its potential for malignancy prediction.

Author Contributions Camilo González, Juan A. Olmos, Catalina de Valencia, and Fabio Martínez contributed equally to the design of the study, implementation of the methods, analysis of the results, and writing of the manuscript.

Funding Open Access funding provided by Colombia Consortium This work was supported by the Vicerrectoría de Investigación y Extensión (VIE) of the Universidad Industrial de Santander, project: *Cuantificación de lesiones de próstata comparando secuencias multiparamétricas y bi-paramétricas de MRI usando herramientas de inteligencia artificial*. Code 3946.

Data availability No datasets were generated or analysed during the current study.

Materials availability All materials and code used in this study will be available upon acceptance.

Declarations

Conflicts of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Sung, H., Ferlay, J., Siegel, R.L.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **71**(3), 209–249 (2021)
2. Futterer, J.J., Briganti, A., De Visschere, P., et al.: Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? a systematic review of the literature. Eur. Urol. **68**(6), 1045–1053 (2015)
3. Merriel, S.W.D., Pocock, L., Gilbert, E.: Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (psa) for the detection of prostate cancer in symptomatic patients. BMC Med. 20(1), (2022)
4. Naji, L., Randhawa, H., Sohani, Z., et al.: Digital rectal examination for prostate cancer screening in primary care: A systematic review and meta-analysis. The Annals of Family Medicine **16**(2), 149–154 (2018)
5. Deniffel, D., Healy, G.M., Dong, X., et al.: Avoiding unnecessary biopsy: Mri-based risk models versus a pi-rads and psa density strategy for clinically significant prostate cancer. Radiology **300**(2), 369–379 (2021)
6. Stabile, A., Giganti, F., Rosenkrantz, A.B., et al.: Multiparametric mri for prostate cancer diagnosis: current status and future directions. Nat. Rev. Urol. **17**(1), 41–61 (2019)
7. Schoots, I.G., Barentsz, J.O., Bittencourt, L.K., et al.: Pi-rads committee position on mri without contrast medium in biopsy-

naive men with suspected prostate cancer: Narrative review. Am. J. Roentgenol. **216**(1), 3–19 (2021)
8. Obmann, V.C., Pahwa, S., Tabayayong, W., et al.: Diagnostic accuracy of a rapid biparametric mri protocol for detection of histologically proven prostate cancer. Urology **122**, 133–138 (2018)
9. Kang, H.C., Jo, N., Bamashmos, A.S., et al.: Accuracy of prostate magnetic resonance imaging: Reader experience matters. European Urology Open Science **27**, 53–60 (2021)
10. Hamdy, F.C., Donovan, J.L., Lane, J.A., et al.: Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. N. Engl. J. Med. **388**(17), 1547–1558 (2023)
11. Bratt, O., Auvinen, A., Godtman, R.A., et al.: Screening for prostate cancer: evidence, ongoing trials, policies and knowledge gaps. BMJ Oncology **2**(1), 000039 (2023)
12. Saha, A., Twilt, J., Joeran, S., et al.: The PI-CAI Challenge: Public Training and Development Dataset. Zenodo (2022)
13. Lay, N., Tsehay, Y., Greer, M.D., et al.: Detection of prostate cancer in multiparametric mri using random forest with instance weighting. Journal of Medical Imaging **4**(2), 024506 (2017)
14. Wang, Z., Liu, C., Cheng, D., et al.: Automated detection of clinically significant prostate cancer in mp-mri images based on an end-to-end deep neural network. IEEE Trans. Med. Imaging **37**(5), 1127–1139 (2018)
15. Yuan, Y., Ahn, E., Feng, D., et al.: Z-SSMNet: Zonal-aware Self-supervised Mesh Network for Prostate Cancer Detection and Diagnosis with Bi-parametric MRI. arXiv (2022). https://arxiv.org/abs/2212.05808
16. Saha, A., Hosseinzadeh, M., Huisman, H., et al.: End-to-end prostate cancer detection in bpmri via 3d cnns: Effects of attention mechanisms, clinical priori and decoupled false positive reduction. Med. Image Anal. **73**, 102155 (2021)
17. Ishioka, J., Matsuoka, Y., Uehara, S., et al.: Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm. BJU Int. **122**(3), 411–417 (2018)
18. Karagoz, A., Seker, M.E., Yergin, M., et al.: Prostate Lesion Estimation using Prostate Masks from Biparametric MRI. arXiv (2023). https://arxiv.org/abs/2301.09673
19. Kan, H., Shi, J., Zhao, M.: Itunet: Integration of transformers and unet for organs-at-risk segmentation. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2123–2127. (2022)
20. Arshad, M., Wang, C., Wajeeh Us Sima, M.: Ranet: a residual attention network for accurate prostate segmentation in t2-weighted mri. Front. Med. 12, (2025). https://doi.org/10.3389/fmed.2025.1589707
21. Arshad, M., Wang, C., Us Sima, M.W., et al.: Bioaug-net: a bioimage sensor-driven attention-augmented segmentation framework with physiological coupling for early prostate cancer detection in t2-weighted mri. BioData Mining 18(1) (2025) https://doi.org/10.1186/s13040-025-00467-4
22. Us Sima, W., Wang, M., Arshad, C., M.: Leveraging advanced feature extraction for improved kidney biopsy segmentation. Front. Med. 12, (2025). https://doi.org/10.3389/fmed.2025.1591999
23. Arshad, M., Wang, C., Ghodhbani, R., et al.: Prostai-net: A federated privacy-preserving deep learning framework for sustainable prostate cancer image classification in digital pathology. International Journal of Computational Intelligence Systems (2026). https://doi.org/10.1007/s44196-026-01271-6
24. Salman, M.E., çakar, G.c., Azimjonov, J., et al.: Automated prostate cancer grading and diagnosis system using deep learning-based yolo object detection algorithm. Expert Syst. Appl. **201**, 117148 (2022)
25. Ou, Y., Yang, R., Ma, L., et al.: Uniinst: Unique representation for end-to-end instance segmentation. Neurocomputing **514**, 551–562 (2022)
26. Liu, S., Qi, L., Qin, H., et al.: Path Aggregation Network for Instance Segmentation. arXiv (2018). https://arxiv.org/abs/1803.01534
27. Bolya, D., Zhou, C., Xiao, F.: YOLACT: Real-time Instance Segmentation. arXiv, (2019)
28. Lin, T.-Y., Maire, M., Belongie, S.J.: Microsoft coco: Common objects in context. CoRR, (2014)
29. Li, X., Vesal, S., Saunders, S.: The prostate imaging: Cancer ai (pi-cai) 2022 grand challenge (pimed team), (2022)
30. Turkbey, B., Rosenkrantz, A.B., Haider, M.A.: Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. Eur. Urol. **76**(3), 340–351 (2019)
31. Saha, A., Twilt, J.J., Bosma, J.S.: Artificial intelligence and radiologists at prostate cancer detection in mri: The pi-cai challenge. Zenodo , (2022)
32. Yan, J., Yu, C., Chen, H.: Seeking common ground while reserving differences: Multiple anatomy collaborative framework for undersampled mri reconstruction. IEEE J. Biomed. Health Inform. , 1–14 (2025)
33. Yu, J., Fulcher, A.S., Winks, S.G., et al.: Diagnosis of typical and atypical transition zone prostate cancer and its mimics at multiparametric prostate mri. Br. J. Radiol. **90**(1073), 20160693 (2017). https://doi.org/10.1259/bjr.20160693
34. Zhou, L., Chen, Z., Shen, Y.: Ersr: An ellipse-constrained pseudo-label refinement and symmetric regularization framework for semi-supervised fetal head segmentation in ultrasound images. IEEE J. Biomed. Health Inform. , 1–11 (2025). https://doi.org/10.1109/jbhi.2025.3602655
35. Jin, Q., Cui, H., Wang, J.: Iterative pseudo-labeling based adaptive copy-paste supervision for semi-supervised tumor segmentation. Knowl.-Based Syst. **324**, 113785 (2025). https://doi.org/10.1016/j.knosys.2025.113785
36. Jin, Q., Cui, H., Sun, C., et al.: Inter- and intra-uncertainty based feature aggregation model for semi-supervised histopathology image segmentation. Expert Syst. Appl. **238**, 122093 (2024). https://doi.org/10.1016/j.eswa.2023.122093

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.