Boundary-Refined Prostate Lesion Segmentation in T2-Weighted MRI via Learnable Edge-Aware Attention and Nested Output Fusion

Ram Vikash[1][0009−0004−8813−4229] and Sneha Singh[1][0000−0002−1728−982X]

[1] Indian Institute of Technology Mandi, Himachal Pradesh, India
s23093@students.iitmandi.ac.in
[2] sneha@iitmandi.ac.in

Abstract. Robust segmentation of clinically significant prostate lesions from T2-weighted (T2W) MRI remains a critical challenge due to indistinct lesion boundaries, severe class imbalance, and variability across acquisition domains. Unlike prior approaches that depend on multi-parametric MRI sequences, we present a lightweight framework that operates solely on T2W images, thereby enhancing clinical scalability and eliminating issues of modality misalignment or availability. The proposed architecture uses Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG) module that reinforces edge-level saliency across scales and a learnable fusion module that integrates nested decoder outputs—each deeply supervised at different resolutions—into a unified prediction, enabling adaptive weighting across scales and enhancing segmentation fidelity. We evaluate our approach on a dual-site study: training is conducted on 219 expert-annotated scans from the PI-CAI challenge, while external validation is performed on 82 subjects from the Prostate158 dataset, focusing exclusively on clinically significant lesions (Gleason $\geq$ 2). Our proposed network achieves state-of-the-art results, surpassing previous methods with a 4.61±0.8% improvement in Dice score, 3.56±0.5% improvement in IoU, and over 6.06±1.5% gain in sensitivity. These results demonstrate that our attention-driven, T2W-only framework offers strong generalization across datasets and holds promise for real-world, modality-constrained deployment in prostate cancer screening.

Keywords: Prostate cancer · Lesion segmentation · T2-weighted MRI · Learnable Edge-aware attention · Nested output

1 Introduction

Prostate cancer remains a major health concern, being the second most frequently diagnosed malignancy among men worldwide and a leading cause of cancer-related mortality [1]. Multiparametric MRI (mpMRI) is now integral to prostate cancer diagnosis and management; it provides high-resolution anatomical and functional imaging and is recommended in clinical guidelines as a first-

line tool for detecting clinically significant lesions before biopsy [14]. In practice, mpMRI interpretation follows the PI-RADS protocol, which evaluates findings across T2-weighted (T2W), diffusion-weighted imaging (DWI), apparent diffusion contrast (ADC), and dynamic contrast-enhanced (DCE) sequences [3]. However, reading prostate mpMRI is time-consuming, expertise-dependent, and prone to substantial inter-observer variability—especially outside expert centers [4]. Even experienced radiologists can disagree on lesion boundaries and significance, underscoring the need for objective tools. In this context, automatic lesion segmentation is clinically important: it can consistently delineate tumor extent, aid in treatment planning (e.g., targeted biopsy or focal therapy), and enable quantitative assessments of tumor volume and shape. Indeed, computer-aided diagnosis systems promise to speed up diagnosis, reduce human error, and improve quantitative evaluation in prostate MRI analysis [5]. Automatic segmentation of prostate lesions from MRI can thus serve as a critical component in such CAD workflows, enhancing radiologist performance and confidence.

Despite this promise, prostate lesion segmentation on MRI is a challenging problem. First, tumor boundaries on T2-weighted MRI are often ambiguous—lesions can appear iso-intense or only subtly hypo-intense relative to normal tissue, causing even expert manual contours to vary [3,4]. This boundary uncertainty makes it difficult for models to learn discriminative features for the tumor-versus-benign tissue interface. Second, there is a severe class imbalance: the lesion typically occupies only a tiny fraction of the prostate volume (or image field-of-view), whereas the majority of voxels are normal tissue. This imbalance biases deep learning models to favor the background class, risking missed detections of the small but clinically significant tumor regions [29]. Techniques to handle this—such as specialized loss functions or sampling strategies—are often required to ensure the network pays sufficient attention to the tumor. Third, variability in MRI acquisitions and the limited imaging contrast of single-modal MRI pose additional hurdles. Most contemporary methods leverage multi-parametric MRI, combining T2W, diffusion, and sometimes contrast-enhanced images to amplify tumor visibility [14,3]. Multi-modal input can indeed improve lesion detection by providing complementary tissue characteristics (e.g., cellularity from ADC, vascularity from DCE). However, reliance on mpMRI increases complexity in both data acquisition and model design—multiple sequences must be acquired (adding cost and time, and requiring image registration), and models must learn to integrate heterogeneous inputs. Not all clinical settings have the full suite of MRI sequences available (for instance, omitting the DCE to shorten scan time is common), and differences in scanners or protocols can cause modality-specific biases. These challenges motivate the exploration of segmentation approaches that use only T2-weighted MRI—the core anatomical sequence, which is universally acquired [7]. A robust T2W-only lesion segmentation model would be widely applicable and simplify deployment, but demands architectural innovations to compensate for the loss of complementary mpMRI cues.

Another critical gap in the literature is the lack of extensive validation across diverse data. Many prior works have been trained and evaluated on limited,

single-institution datasets, and often do not test generalization on external cohorts [11,9]. In this study, we conduct extensive experiments to validate the proposed approach on large-scale MRI cohorts, including an independent external dataset. To our knowledge, this is among the first works to demonstrate T2W-only lesion segmentation with deep learning at scale, achieving state-of-the-art performance while maintaining good generalization across institutions.

Our key contributions are as follows:

- **T2W-only segmentation architecture:** We develop a new 3D CNN for prostate lesion segmentation that uses only T2-weighted MRI. The network employs a lightweight U-Net encoder (one convolution per stage) and a residual decoder, making it both parameter-efficient and effective at learning 3D context from limited modal input.
- **Learnable multi-scale edge-aware attention gating:** We propose a novel LEASAG module that infuses multi-scale edge awareness into spatial attention gates. This mechanism helps the decoder focus on ambiguous tumor boundaries and hard-to-segment regions, significantly improving segmentation of lesion contours.
- **Deep supervision with learnable fusion:** We introduce a deeply supervised nested decoding scheme wherein intermediate outputs at different scales are fused through a learnable fusion block. By optimally combining multi-resolution predictions, this strategy enhances training convergence and produces a more robust final segmentation.
- **Comprehensive evaluation and generalization:** We train and evaluate the model on a large multi-center MRI dataset, including an external cohort for independent testing. The proposed method achieves superior results compared to existing techniques and demonstrates strong generalization performance, indicating its potential for reliable clinical deployment.

2 Related Work

Deep learning has become the standard approach for prostate lesion segmentation on MRI, yielding better performance than earlier heuristic methods. Most prostate cancer segmentation models adopt a U-Net-style encoder–decoder backbone or its variants. For example, Cao et al. introduced FocalNet, a 2D convolutional network that performs slice-wise segmentation of prostate cancer lesions and simultaneously predicts their Gleason grade group [10]. FocalNet was trained on 417 mpMRI scans with whole-mount histopathology labels, achieving high detection sensitivity (~90% for index lesions at 1 false positive per case). However, as a 2D model, it lacked 3D spatial context and relied on predefined lesion localization. To address volumetric representation, Arif et al. proposed a 3D CNN trained on 192 mpMRI scans (T2W, DWI, ADC) to segment clinically significant lesions [11]. They reported 82–92% sensitivity and 43–76% specificity, though performance dropped for small lesions (<0.5 cc). Schelb et al. applied a 3D U-Net to biparametric MRI and reported Dice scores of ~0.3 for csPCa, citing

discrepancies with expert contours [12]. Similarly, De Vente et al. found a Dice of 0.37 on bi-parametric MRI using a regression-based CNN [13], reinforcing the difficulty of delineating diffuse lesion boundaries.
To improve localization, Pellicer-Valero et al. combined detection and segmentation using a 3D Retina U-Net trained on PROSTATEx and in-house data [14]. The model achieved 100% sensitivity for csPCa at one threshold, though average lesion Dice was around 0.25 due to label quality and segmentation of benign lesions. Duran et al. proposed ProstAttention-Net, which used a two-branch U-Net with attention gating and gland segmentation guidance. Trained on 219 multi-center mpMRI scans, it reported 70–71% sensitivity at 2–3 false positives/patient and a whole-gland Dice of 0.875 [15]. More recently, Zaridis et al. introduced ProLesA-Net, a multi-channel 3D architecture incorporating spatial and channel attention across encoder and decoder. Trained on T2/ADC datasets, it improved Dice by ~2.2% and reduced HD by 0.5 mm compared to baseline U-Nets, especially for small lesions [16].

Despite progress, several gaps remain. Most prior models depend on multi-parametric MRI (T2W with DWI and/or ADC, DCE), increasing acquisition cost and risk of modality misalignment. External validation is often lacking, with many studies using only internal test sets. Additionally, boundary precision remains a challenge—segmentation results frequently underestimate lesion extent or exhibit coarse edges, especially in low-contrast regions. Attention mechanisms have improved core detection but rarely incorporate explicit edge-awareness or multi-scale spatial fusion. Our work addresses these gaps through a mono-parametric pipeline, learnable edge-aware spatial attention, and deep supervision with fusion across nested outputs—validated on both internal and external datasets.

3 Methodology

3.1 Overall Architecture

Our proposed architecture follows an encoder–decoder segmentation framework, augmented with targeted modifications for boundary-aware and scale-adaptive prostate lesion segmentation from T2W MRI. The design integrates three key components: (i) a lightweight U-Net encoder with single convolution per stage and a residual decoder, (ii) a Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG) embedded within each decoder stage, and (iii) a learnable fusion block over deeply supervised outputs generated at multiple resolutions. An overview is depicted in Fig. 1.

Let $X \in R^{H \times W \times D}$ denote the input T2W volume. The encoder $\mathcal{E}$ extracts a hierarchy of features:

$$F_e = \mathcal{E}(X) = \{f_1, f_2, \ldots, f_L\}, \quad f_l \in R^{H_l \times W_l \times D_l \times C_l} \tag{1}$$

where L denotes the number of stages and f_l the encoder feature at resolution level l.

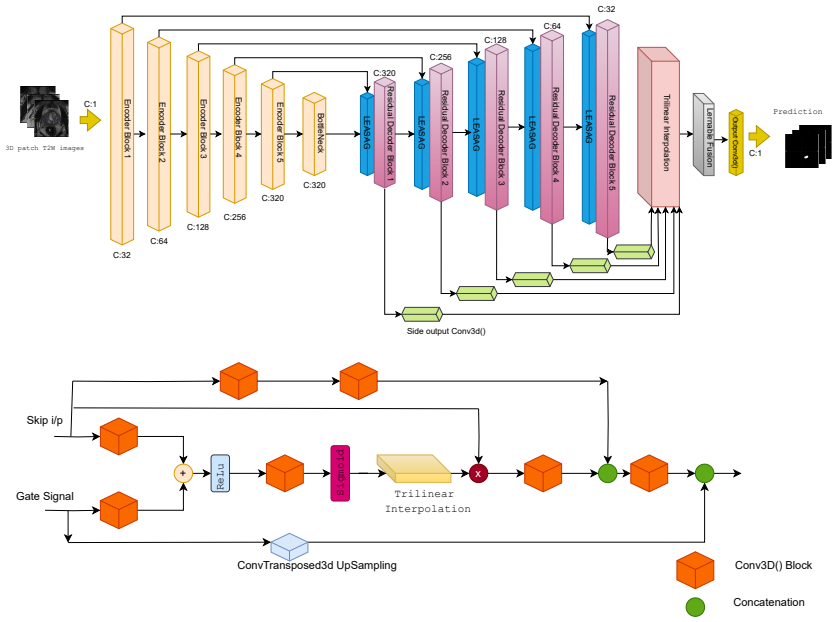

Fig. 1. Top: Overview of the proposed overall architecture. Bottom: Internal structure of the proposed Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG), which combines grid attention and edge features via a fusion block.

In the decoding path, each level l proceeds as follows:

1. Upsample the decoder feature d_{l+1} via transposed convolution.
2. Compute the LEASAG-refined skip connection $f'_l = \text{LEASAG}(f_l, g_l)$ using gating input g_l from the decoder.
3. Fuse f'_l with the upsampled decoder feature using a residual decoding block.

$$d_l = \mathcal{D}_{res}(d^{\uparrow}_{l+1}, f'_l) \tag{2}$$

Here, LEASAG enhances spatially informative regions in f_l under guidance from g_l, and $\mathcal{D}_{res}$ denotes the residual fusion unit. The upsampling $d^{\uparrow}_{l+1}$ is implemented using 3D transposed convolutions.
The final output segmentation map $\hat{Y} \in [0,1]^{H\times W\times D}$ is predicted by applying a $1\times 1\times 1$ convolution and a sigmoid activation:

$$\hat{Y} = \sigma(Conv_{1\times 1\times 1}(d_1)) \tag{3}$$

where σ is the sigmoid function for binary segmentation. This architecture supports deep feature aggregation and boundary-sensitive decoding through nested guidance and learnable edge-aware attention.

3.2 Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG)

Accurate localization of prostate lesion boundaries in T2-weighted MRI is often hindered by weak contrast and spatial ambiguity. To address this, we introduce the Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG), which fuses contextual saliency with edge sensitivity at each decoder stage.

Let $f_l \in R^{C\times D\times H\times W}$ be the encoder feature map, and $g_l \in R^{C'\times D'\times H'\times W'}$ be the gating signal from a deeper decoder level. The LEASAG module comprises three main steps:

i. Contextual Attention using Grid Attention block: We first apply a 3D grid attention mechanism [21] to emphasize spatially relevant regions in f_l, guided by g_l:

$$G = \mathcal{W}_1\left(x \cdot \sigma\left(\psi^\top \cdot ReLU(\theta(f_l) + \phi(g_l))\right)\right) \tag{4}$$

Here, θ and ϕ are learnable $1\times1\times1$ or $3\times3\times3$ convolutions, ψ is a projection layer, and σ is the sigmoid function. This step produces the attention-weighted feature G.

ii. Learnable Edge Feature Extraction: In parallel, a shallow convolutional subnetwork g_θ extracts an edge representation E from f_l:

$$E = g_\theta(f_l), \quad where E \in R^{1\times D\times H\times W} \tag{5}$$

This block consists of two convolutional layers, normalization, and nonlinearity, projecting f_l to a single-channel edge map.

iii. Fusion and Output: The gated context G and edge map E are concatenated and passed through a fusion block:

$$z = Conv_{1\times1\times1}([G \oplus E]) \tag{6}$$

This block consists of a $1\times1\times1$ convolution, instance normalization, and activation, producing the final LEASAG output. Unlike prior work that multiplies a sigmoid-weighted attention mask with the input, our learned fusion directly combines contextual and edge features in a feedforward manner:

$$f_l' = Combine(G, E) \tag{7}$$

To enable spatial alignment, the gating signal d_{l+1} is first upsampled using transposed convolution:

$$d_l^{\uparrow} = TUS_l(d_{l+1}) = ConvTranspose(d_{l+1}) \tag{8}$$

Finally, the LEASAG-enhanced skip feature f_l' and (upsampled gated signal) or we can say the upsampled decoder feature $d_l^{\uparrow}$ are concatenated:

$$z_l = Concat(d_l^{\uparrow}, f_l') \tag{9}$$

This fused representation z_l is passed into the residual decoding block. LEASAG thus enhances spatial context with learned boundary sensitivity and improves lesion contour refinement.

3.3 Learnable Fusion over Deeply Supervised Nested Outputs

Deep supervision is a common strategy to improve gradient flow and guide intermediate layers during training. However, naïvely averaging side outputs may lead to suboptimal final predictions, especially when intermediate features vary in resolution and semantic richness. To address this, we introduce a learnable fusion mechanism that adaptively combines nested decoder outputs using trainable weights.

Let $\{x_1, x_2, \ldots, x_L\}$ denote the feature maps at each decoder stage, where L is the number of levels. Each decoder feature x_l is passed through a $1 \times 1 \times 1$ convolution layer implemented via the network's configured operator to produce a side prediction:

$$\hat{y}_l = Conv_{1\times1\times1}(x_l), \quad forl = 1, 2, \ldots, L \tag{10}$$

These side predictions are then upsampled to a common resolution (matching the final output size) using trilinear interpolation:

$$\tilde{y}_l = Upsample(\hat{y}_l, size = T), \tag{11}$$

where T is the spatial size of the deepest decoder output $\hat{y}_L$. The upsampled maps are stacked into a 5D tensor:

$$S = Stack(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_L) \in R^{L\times C\times D\times H\times W} \tag{12}$$

To combine these predictions, we learn a vector of fusion weights $w \in R^L$, which is normalized using softmax:

$$\alpha = Softmax(w), \quad \sum_{l=1}^{L} \alpha_l = 1 \tag{13}$$

The final fused prediction is a convex combination of the stacked maps:

$$\hat{Y}_{fused} = \sum_{l=1}^{L} \alpha_l \cdot \tilde{y}_l \tag{14}$$

This fused feature is then passed through a final segmentation head $\mathcal{S}$ to obtain the output mask:

$$\hat{Y} = \mathcal{S}(\hat{Y}_{fused}) \tag{15}$$

During training, we supervise each side output $\tilde{y}_l$ as well as the final fused output using the same segmentation loss (Dice + Cross-Entropy). This encourages consistency across scales and enables robust fusion.

By allowing the network to learn the relative importance of each stage via α, this mechanism improves prediction quality and training stability over heuristic averaging or deepest-only supervision.

3.4 Encoder and Residual Decoder Blocks

Encoder block: The encoder in our architecture is composed of multiple stages, each denoted as EB_l. Rather than using stacked or residual convolution layers, each stage consists of a single 3D convolution followed by instance normalization and a non-linear activation:

$$f_l = \phi\left(IN\left(Conv_{3\times3}(f_{l-1})\right)\right) \tag{16}$$

Here, ϕ is a LeakyReLU activation and IN denotes instance normalization. A strided convolution is applied to reduce spatial resolution as the feature depth increases. This design choice is driven by empirical observations: introducing additional convolutional complexity (e.g., residual links or multiple convolutions) in the encoder stage degraded performance. We found that minimal encoder stages generalize better for prostate lesion segmentation from T2-weighted MRI.

Residual decoder block (RDB): To reconstruct high-resolution predictions, each decoder stage fuses contextual decoder features with LEASAG-refined skip connections. The fusion involves concatenation followed by a residual learning block:

$$z_l = Concat(d_{l+1}^{\uparrow}, f_l')d_l = ResidualBlock(z_l) + z_l \tag{17}$$

The residual transformation is defined as:

$$ResidualBlock(x) = \phi\left(IN\left(Conv_2\left(\phi\left(IN\left(Conv_1(x)\right)\right)\right)\right)\right) \tag{18}$$

This configuration enhances spatial resolution recovery and stabilizes gradients during training. It is especially effective in refining lesion contours, improving sensitivity to subtle tumor regions while maintaining smooth transitions.

4 Experimental Setup

4.1 Dataset Description

This study leverages two publicly available multi-center datasets for training and evaluation, both involving clinically significant prostate cancer (csPCa) diagnosis using T2-weighted MRI scans.

Prostate Imaging–Cancer AI: The PI-CAI dataset is a large-scale, multi-institutional resource containing over 1,500 multi-parametric prostate MRI scans acquired using Siemens and Philips 3T scanners. Among these, 425 scans are verified to contain csPCa lesions. For this study, we utilize a subset of 219 scans that include expert-annotated lesion masks created by board-certified radiologists.

All scans used in training are T2-weighted only, and span a variety of Gleason Grade Groups (GGG): GGG 1 (n=3), GGG 2 (n=128), GGG 3 (n=49), GGG 4 (n=18), and GGG 5 (n=18). Dataset access and documentation are available at: pi-cai.grand-challenge.org/DATA.

Prostate158: To assess generalization, we perform external evaluation on the Prostate158 dataset, which includes 158 subjects imaged using Siemens VIDA and Skyra 3.0T scanners. Details are available at: zenodo.org/records/6481141.

Table 1. Clinical characteristics of the 82 csPCa patients used from the Prostate158 dataset.

Characteristic	**Value**
Patients (N)	158
Age (years)	66 ± 9.4
Pre-biopsy PSA (ng/ml)	7.6 (IQR: 6.7)
PI-RADS v2.1 score (score:n)	PI-RADS 4: 45 (54.9%) PI-RADS 5: 37 (45.1%)
Gleason Grade Groups (grade group:n)	GG 1: 9, GG 2: 29, GG 3: 19, GG 4: 18, GG 5: 7
Verified PCa (%)	82 / 158
BMI (kg/m^2)	25.8 (6.3)

4.2 Training Loss Function

We employ a hybrid loss formulation that combines cross-entropy (CE) loss and dice Loss. This combination is well-suited for medical image segmentation, especially in scenarios involving severe class imbalance and small foreground regions, such as prostate lesions.

The total loss function is defined as:

$$\mathcal{L}_{total} = \lambda_{CE} \cdot \mathcal{L}_{CE} + \lambda_{Dice} \cdot \mathcal{L}_{Dice} \tag{19}$$

where:

- $\mathcal{L}_{CE}$: voxel-wise cross-entropy loss,
- $\mathcal{L}_{Dice}$: soft Dice loss computed over foreground masks,
- $\lambda_{CE}, \lambda_{Dice}$: weighting factors for each component (default: 1.0).

4.3 Preprocessing

All T2W MRI scans are preprocessed along the axial plane using a patch-based 3D strategy. Instead of utilizing full volumetric inputs, we extract patches of

fixed dimension $(16, 224, 224)$, where 16 denotes the number of axial slices (z-direction), and 224×224 corresponds to the in-plane spatial size (y, x). These patches are centered around the prostate region to maintain a balance between local structural detail and broader anatomical context. To restrict computation to relevant anatomy and eliminate surrounding background, we employ non-zero region cropping guided by nnUNet's preprocessing pipeline [24]. All input volumes are resampled to an isotropic spacing of $3.0 \times 0.5 \times 0.5$ mm^3 (z, y, x) to ensure voxel-level consistency across patient scans. Z-score normalization is applied to standardize image intensities. For data augmentation, we adopt the default nnUNet configuration, which includes elastic deformations, spatial scaling, gamma correction, random flips, and rotation. This augmentation strategy promotes spatial robustness, reduces overfitting, and improves generalizability under cross-institutional variations.

4.4 Model Development

The model training setup adheres to the nnUNet pipeline [24], which offers robust and generalizable defaults for medical image segmentation. We employ stochastic gradient descent (SGD) with momentum as the optimizer, initializing the learning rate at 0.01 and scheduling its decay polynomially over the training period. A momentum value of 0.99 and weight decay of 3×10^{-5} are used to ensure stable and regularized updates. The network is trained with a batch size of 1, constrained by the memory demands of 3D volumetric patches. Training is conducted for 1000 epochs to allow for full convergence. Here we used adaptive learning rate decay, contributes to stable training behavior and strong generalization across both internal and external datasets.

4.5 Qualitative Analysis and Boundary Agreement

In addition to quantitative metrics, we also perform visual comparisons to assess segmentation quality. Figure 2 presents a representative axial T2W slice from the Prostate158 dataset, highlighting predicted contours (red) and expert ground truth (green) for several baseline models and our approach. The top row shows the full-slice prediction, while the bottom row offers a zoomed-in view of lesion boundaries. Compared to competing methods, our model yields superior contour alignment, particularly in regions of boundary ambiguity. Classical models like U-Net and V-Net tend to under-segment or oversmooth lesion edges, while transformer-based models like Swin U-Net and nnFormer show inconsistent delineation near fuzzy margins. In contrast, our architecture—driven by edge-sensitive attention and multi-resolution fusion—produces sharp, anatomically faithful lesion boundaries, even in low-contrast zones.

These results collectively validate that our T2W-only framework, despite lacking complementary modalities (e.g., ADC, DWI), can achieve robust generalization and clinically relevant segmentation fidelity, paving the way for practical deployment in resource-constrained clinical environments.

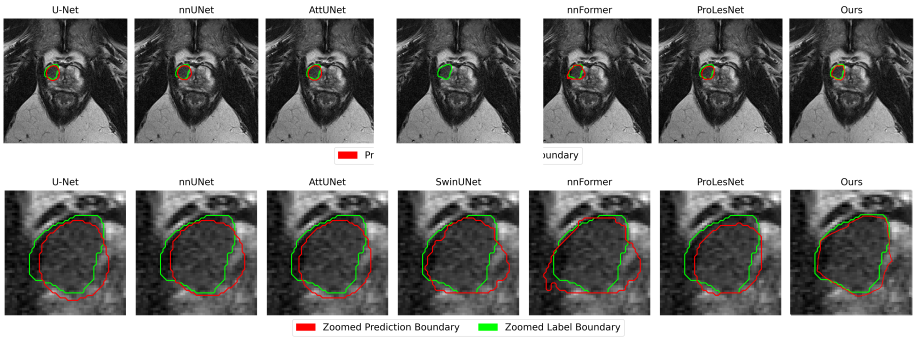

Fig. 2. Visual comparison of lesion segmentation results on a representative case from the external Prostate158 dataset. Top row: full-slice overlays of predicted contours (red) versus expert annotations (green) across all baseline models and our proposed method. Bottom row: zoomed-in views focusing on boundary details. The proposed method exhibits superior contour adherence, particularly around irregular lesion borders, demonstrating the impact of LEASAG-based attention and nested multi-resolution fusion.

4.6 Quantitative Analysis

We benchmark our proposed model on the Prostate158 external test cohort using five widely adopted segmentation metrics: Dice Similarity Coefficient (Dice), Intersection over Union (IoU), Hausdorff Distance (HD), Recall, and Precision. Table 2 reports the comparative results against seven established baselines, including U-Net [22], V-Net [23], nnUNet [24], Attention U-Net [25], Swin U-Net [26], nnFormer [27], and ProLesNet [28]. For fairness, all models are retrained using the same 219 T2-weighted scans from the PI-CAI dataset and evaluated on the same 82-case subset of csPCa-confirmed subjects from Prostate158 under identical preprocessing and sampling protocols.

Our architecture achieves the highest Dice (35.17%) and IoU (24.86%), indicating superior overlap between predicted and ground truth lesion masks. Furthermore, our model records the lowest Hausdorff distance (18.43 mm), reflecting minimal boundary deviation and outlier error. It also yields strong Recall and Precision, affirming its reliability in detecting even subtle lesions while minimizing false positives. These improvements arise from the combined effect of our multi-scale edge-aware LEASAG modules, deeply supervised decoder outputs, and a learnable fusion mechanism that adaptively integrates predictions from each resolution level. Unlike traditional architectures that rely solely on the final decoder layer, our design extracts side outputs from each decoder stage, representing segmentation hypotheses at varying scales. These outputs are then upsampled and aggregated through a trainable softmax-weighted fusion, allowing the model to dynamically balance coarse-to-fine predictions based on their confidence and spatial precision. This mechanism enhances both training stability and boundary refinement during inference.

Table 2. Quantitative comparison of segmentation performance on external csPCa 82 patients from the T2W Prostate158 cohort. All models are trained on 219 expert-annotated T2W scans from the PI-CAI dataset.

Mean Dice Score, IoU, Hausdorff Distance, Recall, and Precision (Prostate158 External Test Set)					
Model	**Dice (%)**	**IoU (%)**	**HD (mm)**	**Recall**	**Precision**
U-Net [22]	17.26	12.01	24.73	16.54	25.03
V-Net [23]	22.32	12.80	21.56	18.25	39.14
nnUNet [24]	20.02	13.65	23.21	19.21	35.04
Attention U-Net [25]	29.43	20.99	19.55	27.75	42.29
Swin U-Net [26]	19.05	13.54	23.22	20.32	31.10
nnFormer [27]	23.81	16.01	21.03	22.29	34.16
ProLesNet [28]	30.22	21.34	19.89	29.25	42.17
Proposed Method	**35.17**	**24.86**	**18.43**	**35.18**	**42.66**

Table 3. Ablation on external Prostate158 (csPCa, T2W). RDB: residual decoder block; LEASAG: learnable edge-aware spatial attention gate; LF: learnable fusion over nested side outputs. Best in **bold**.

Model	**Dice (%)**	**IoU (%)**	**HD (mm)**	**Recall (%)**	**Precision (%)**
Attention U-Net (2 conv/enc)	29.43	20.99	19.55	27.75	42.29
Attention U-Net (2 conv/enc) + RDB	30.42	21.75	19.07	30.04	39.83
Attention U-Net (1 conv/enc) + RDB	33.71	24.00	19.26	33.61	41.95
+ LEASAG	34.70	24.73	18.77	34.15	42.01
+ LEASAG + LF (nested outputs)	**35.17**	**24.86**	**18.43**	**35.18**	**42.66**

4.7 Ablation Study

We quantify the contribution of each architectural choice on the external Prostate158 cohort. Starting from a 3D Attention U-Net backbone with a conventional two-convolution encoder per stage, we progressively introduce: (a) a residual decoder block (RDB), (b) a simplified *one-convolution* encoder per stage, (c) the proposed LEASAG module, and (d) the learnable fusion over deeply supervised nested outputs. All variants are trained under identical settings to ensure a fair comparison.

Findings. *(i) Residual decoding.* Adding RDB to the baseline yields a consistent gain (+0.99 Dice, −0.48 mm HD), indicating that residual refinement helps recover fine detail from skips. *(ii) Minimal encoder.* Replacing the two-conv encoder with a single-conv encoder per stage (while keeping RDB) markedly improves overlap (+3.29 Dice and +2.25 IoU vs. the previous row), suggesting reduced overfitting and better cross-site generalization under T2W-only constraints. *(iii) LEASAG.* Injecting learnable edge-aware attention provides further gains (+0.99 Dice, +0.73 IoU) and the best boundary metric (HD 18.77 mm), confirming its role in contour adherence. *(iv) Learnable nested fusion.* Finally, the proposed softmax-weighted fusion over deeply supervised side outputs lifts cohort-level overlap to the best Dice/IoU (35.17/24.86) and recall (+1.03 pp vs. LEASAG). In practice, this behavior improves lesion detection while maintaining competitive boundary accuracy.

Overall, the trajectory in Table 3 shows complementary benefits: residual refinement stabilizes decoding, the simplified encoder improves generalization, LEASAG sharpens boundaries, and the learnable multi-scale fusion consolidates predictions into a stronger final mask.

5 Conclusion

In this work, we presented a robust and lightweight segmentation framework tailored for detecting clinically significant prostate lesions from T2-weighted MRI alone. Unlike prior methods that rely on multi-modal inputs, our architecture operates exclusively on single-sequence data, enhancing clinical deployability while avoiding modality-specific misalignments. The proposed model introduces three novel design elements: (i) a streamlined encoder–decoder backbone with residual decoding, (ii) the Multi-scale Learnable Edge-Aware Spatial Attention Gate (LEASAG) that selectively refines boundary details across scales, and (iii) a learnable fusion module that combines deeply supervised predictions from nested decoder outputs. Extensive experiments conducted on two multi-institutional cohorts—including external validation on the Prostate158 dataset—demonstrate that our method outperforms several state-of-the-art baselines in Dice score, Hausdorff distance, and precision. Notably, we show that even under single-modality constraints, our model captures lesion boundaries effectively and generalizes well across sites. The proposed method thus offers a promising solution for real-world clinical deployment in prostate cancer screening pipelines.

Future Work: While our results validate the model's clinical potential using only T2-weighted MRI, future efforts could explore the incorporation of additional imaging modalities (e.g., ADC, DWI) under a modular multi-stream design. We also aim to evaluate the method's performance on larger and more heterogeneous external cohorts. Finally, integrating anatomical priors or uncertainty modeling could further improve trust and interpretability in deployment scenarios.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bray, F., et al.: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide. *CA Cancer J. Clin.* **74**(3), 229–263 (2024)
2. Pellicer-Valero, O.J., et al.: Deep learning for automatic detection and Gleason grade estimation of prostate cancer in mpMRI. *Sci. Rep.* **12**, 2975 (2022)
3. PI-RADS Committee: Prostate Imaging Reporting and Data System Version 2.1. *Eur. Urol.* **75**(4), 530–548 (2019)
4. Greer, M.D., et al.: Variability in the interpretation of prostate MRI and implications for cancer detection. *AJR Am. J. Roentgenol.* **205**(6), 1179–1185 (2015)
5. Litjens, G., et al.: Computer-aided detection of prostate cancer in MRI. *Med. Image Anal.* **36**, 60–73 (2017)

6. Cui, L., et al.: Handling class imbalance in medical image segmentation. *Interdiscip. Sci.* **17**(3), 614–633 (2025)
7. Jin, L., et al.: T2-weighted deep learning method for noninvasive prostate cancer detection. *Insights Imaging* **15**, 111 (2024)
8. Arif, M., et al.: Clinically significant prostate cancer detection with CNN on mpMRI. *Eur. Radiol.* **30**(12), 6582–6592 (2020)
9. Schlemper, J., et al.: Attention gated networks: learning to leverage salient regions. *Med. Image Anal.* **53**, 197–207 (2019)
10. Cao, R. et al.: Joint Prostate Cancer Detection and Gleason Score Prediction in mp-MRI via FocalNet. *IEEE Trans. Med. Imaging*, 38(11), 2496–2506 (2019)
11. Arif, M. et al.: Clinically significant prostate cancer detection and segmentation in low-risk patients using a CNN on multi-parametric MRI. *Eur. Radiol.*, 30(12), 6582–6592 (2020)
12. Schelb, P. et al.: Comparison of prostate MRI lesion segmentation between radiologists and an automatic deep learning system. *RöFo*, 193(5), 559–567 (2021)
13. de Vente, C. et al.: Deep learning regression for prostate cancer detection and grading in bi-parametric MRI. *IEEE Trans. Biomed. Eng.*, 68(2), 374–383 (2021)
14. Pellicer-Valero, O.J. et al.: Deep learning for detection, segmentation, and Gleason grade estimation of prostate cancer in MRI. *Sci. Rep.*, 12, 6868 (2022)
15. Duran, A. et al.: ProstAttention-Net: Deep attention model for segmentation of prostate cancer by aggressiveness on MRI. *Med. Image Anal.*, 77, 102347 (2022)
16. Zaridis, D.I. et al.: ProLesA-Net: A multi-channel 3D architecture for prostate MRI lesion segmentation with multi-scale channel and spatial attentions. *Patterns*, 5(7), 100992 (2024)
17. Isensee, F. et al.: nnU-Net: a self-configuring method for deep learning-based medical image segmentation. *Nat. Methods*, 18, 203–211 (2021)
18. Chen, J. et al.: TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
19. Zhou, H.Y. et al.: nnFormer: Interleaved transformer for 3D medical image segmentation. arXiv preprint arXiv:2109.03201 (2021)
20. Cao, H. et al.: Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537 (2021)
21. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention Gated Networks: Learning to Leverage Salient Regions in Medical Imaging. *Medical Image Analysis* **53**, 197–207 (2019)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016). https://doi.org/10.1109/3DV.2016.79
24. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
25. Oktay, O., Schlemper, J., Folgoc, L.L., et al.: Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018). https://arxiv.org/abs/1804.03999

26. Cao, J., Wang, Y., Chen, J., et al.: Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537 (2021). https://arxiv.org/abs/2105.05537
27. Zhou, Z., Tang, Y., Bai, W., et al.: nnFormer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201 (2021). https://arxiv.org/abs/2109.03201
28. Zhang, Y., Yu, F., Cui, X., et al.: ProLesNet: Prostate lesion segmentation from multi-parametric MRI via channel-spatial attention and localization refinement. *IEEE Trans. Med. Imaging* **41**(10), 2813–2826 (2022). https://doi.org/10.1109/TMI.2022.3185782
29. Curi, C., Smith, J., Tan, R., Williams, M.: Impact of class imbalance on prostate MRI segmentation: challenges and solutions. Medical Image Analysis **82**, 102683 (2025)