

# A Second-Order Attention Mechanism For Prostate Cancer Segmentation and Detection in Bi-Parametric MRI

Mateo Ortiz<sup>1</sup>[0009–0006–8343–8084], Juan A. Olmos<sup>1,2</sup>[0000–0002–6017–0867], and  
Fabio Martínez<sup>1</sup>[0000–0001–7353–049X]

<sup>1</sup> Biomedical Imaging, Vision and Learning Laboratory (BIVL<sup>2</sup>ab), Universidad Industrial de Santander (UIS), 680002 Bucaramanga, Colombia  
mateo2201778@correo.uis.edu.co, jaolmosr@correo.uis.edu.co,  
famarcar@saber.uis.edu.co

<sup>2</sup> U2IS, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France.

**Abstract.** The detection of clinically significant prostate cancer lesions (csPCa) from biparametric magnetic resonance imaging (bp-MRI) has emerged as a noninvasive imaging technique for improving accurate diagnosis. Nevertheless, the analysis of such images remains highly dependent on the subjective expert interpretation. Deep learning approaches have been proposed for csPCa lesions detection and segmentation, but they remain limited due to their reliance on extensively annotated datasets. Moreover, the high lesion variability across prostate zones poses additional challenges, even for expert radiologists. This work introduces a second-order geometric attention (SOGA) mechanism that guides a dedicated segmentation network, through skip connections, to detect csPCa lesions. The proposed attention is modeled on the Riemannian manifold, learning from symmetric positive definite (SPD) representations. The proposed mechanism was integrated into standard U-Net and nnU-Net backbones, and was validated on the publicly available PI-CAI dataset, achieving an Average Precision (AP) of 0.37 and an Area Under the ROC Curve (AUC-ROC) of 0.83, outperforming baseline networks and attention-based methods. Furthermore, the approach was evaluated on the Prostate158 dataset as an independent test cohort, achieving an AP of 0.37 and an AUC-ROC of 0.75, confirming robust generalization and suggesting discriminative learned representations.

**Keywords:** Clinically significant prostate cancer lesions · Detection · Segmentation · Geometric Attention · Second-order descriptors

## 1 Introduction

Prostate cancer (PCa) is the second most common cancer and the fifth leading cause of cancer-related death among men. In 2022, over 1.5 million new cases and over 390,000 deaths were reported globally [2]. PCa detection commonly relies on the prostate-specific antigen (PSA) test, but reporting a remarked low

specificity with around 36%, leading to many false positives and unnecessary procedures [16]. Digital rectal examination (DRE) is also used but is invasive, subjective, and limited in regions beyond rectal wall [16].

Recently, multiparametric magnetic resonance imaging (mp-MRI) has demonstrated significant support in pre-biopsy assessment by localizing clinically significant prostate cancer (csPCa) lesions, highlighting both vascular and morphological tissue properties [23]. However, mp-MRI involves long acquisition times and the use of contrast agents. Alternatively, bi-parametric MRI (bp-MRI) excludes dynamic contrast enhancement (DCE) and offers comparable diagnostic accuracy, providing a practical solution for large-scale screening and routine clinical use [5]. Nonetheless, in both cases mp-MRI and bp-MRI, the interpretation and characterization of lesion findings remain subjective and expert-dependent, often leading to confusion between malignant lesions and prostatic hyperplasia [22, 5, 23].

Recently, U-Net based architectures have been mainly proposed for the detection of csPCa in bp-MRI due to their ability to capture multiscale contextual features [18]. Other works have integrated attention modules refining channel- and spatial-level representations during the fusion of encoder and decoder features but depending on large-scale annotated datasets [20, 6]. When trained on limited data or single-center cohorts, a common scenario, attention-based models often show reduced performance and generalization, particularly on external datasets [4].

This work proposes a second-order geometric attention (SOGA) module to capture more relevant bp-MRI deep representations for the detection of csPCa lesions. The *SOGA* mechanism compresses feature banks into compact second-order descriptors that summarize correlations between features. These resulting descriptors are symmetric definite positive (SPD) matrices that coexist in a smooth Riemannian manifold. Consequently, we considered Riemannian deep learning layers to learn SPD geometric descriptors. After this, the resultant SPD representations are projected into a Euclidean space and used for refining feature banks, enabling the network to learn more discriminative patterns from second-order information. The proposed *SOGA* was integrated in a standard U-Net architecture and also embedded within the nnU-Net framework. The inclusion of *SOGA* demonstrated to learn more discriminative and generalizable features, improving detection performance, especially in external cohort evaluation.

## 2 Related Works

Convolutional neural networks (CNNs) have been the standard tool to build computer-aided diagnosis (CAD) systems, dedicated on detecting malignant regions in medical imaging [15]. For csPCa detection, encoder-decoder architectures such as U-Net have become the standard, given their ability to capture malignancy patterns at the pixel level [18]. Nonetheless, csPCa detection remains challenging due to the small size of many lesions, the low contrast in bp-MRI scans, and the presence of anatomically complex regions such as the central zone,

where benign structures may closely resemble malignant tissue [22]. To overcome these limitations, attention mechanisms have been introduced to enhance feature representation. For example, *Wei et al.* proposed an Attention U-Net that refines encoder-to-decoder feature maps and incorporates prostate zone information to anatomically guide the network, thereby improving detection sensitivity [24]. *Yang et al.* applied channel-wise attention using Squeeze-and-Excitation (SE) blocks to enhance the refinement of deep features. However, their approach focused on subregions rather than entire bp-MRI studies, limiting the network’s ability to capture broader contextual anatomical information [26]. *Li et al.* recently introduced a U-Net that also integrated anatomical information, leveraging it through a zone-aware loss function to improve lesion segmentation [13]. *Duran et al.* proposed a dual-decoder network that jointly segments the prostate gland and lesions, using multiplicative spatial attention to infuse anatomical context into lesion segmentation [4].

Despite recent advances in attention-based methods for csPCa detection, their performance shows poor generalization capabilities, *i.e.*, evaluating the approach over unseen external datasets. This motivates the adoption of self-configuring frameworks such as nnU-Net, which autonomously adapts pre-processing, training pipelines, and hyperparameters to the target data [11]. Thus, recent approaches for csPCa detection have increasingly adopted the nnU-Net framework. For instance, Debs et al. demonstrated its superiority over the standard U-Net [3]. This has inspired new methods based on the nnU-Net framework. *Karagoz et al.*, for example, integrated probabilistic prostate zone masks as additional input channels, emphasizing the importance of anatomical context [12]. Although the anatomical context improves lesion detection, many existing methods continue to perform poorly in external data sets [13]. To overcome this, transformer-based architectures like CSwin-UNet leverage self-supervised pretraining through contrastive learning and image restoration tasks, yielding more generalizable feature representations, but with high dependency on large datasets [14].

### 3 Proposed Method

We propose a second-order geometric attention (*SOGA*) module for enhance the encoder’s skip connections by capturing channel-wise and spatial interdependencies. The proposed approach first compresses high-dimensional feature maps into compact symmetric positive definite (SPD) matrices, effectively encoding inter-channel correlations. Then, the method leverages a Riemannian processing pipeline to preserve the intrinsic geometry of the SPD manifold throughout the learning process. Finally, SPD embeddings are mapped to Euclidean space, and linear projections are applied to compute channel-wise weights that recalibrate the features information to be passed to the following layers in the decoder. Figure 1 illustrates the pipeline of the proposed *SOGA* module.

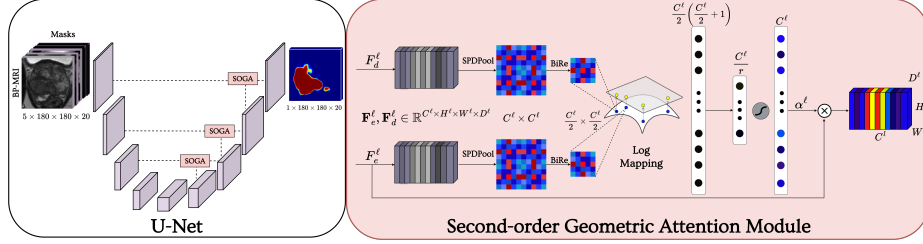


Fig. 1: **Pipeline** of the proposed sencod-order geometric attention (*SOGA*) module integrated in a U-Net network.

### 3.1 U-Net-based representation

This work considers current end-to-end segmentation architectures for the automatic delineation of csPCa lesions on bp-MRI. Among these, the U-Net remains the most widely adopted model for biomedical image segmentation because of its ability to capture multiscale visual context via skip connections between the encoder and decoder [18]. In brief, the U-Net architecture consists of an encoder that at each level  $\ell$  progressively computes blocks of features denoted by  $\mathbf{F}_e^\ell \in \mathbb{R}^{C_\ell \times H_\ell \times W_\ell \times D_\ell}$ , where  $C_\ell$  is the number of feature channels at level  $\ell$ ,  $H_\ell \times W_\ell$  are the spatial dimensions, and  $D_\ell$  is the depth. These layers continue until they converge to an embedded representation (the bottleneck), from which the decoder then reconstructs the lesion delineation by progressively upsampling through decoder feature blocks  $\mathbf{F}_d^\ell$ . At each level  $\ell$ , skip connections transfer encoder information to complement the decoder’s contextual representations, thereby improving segmentation performance.

Recently, skip connection has been enhanced with the incorporation of attention mechanisms, which accentuate the localization of meaningful structures, while progressively suppressing irrelevant background [17, 25]. This process involves recalibrating the feature activations in alignment with the target task. To this end, the input feature maps from the encoder at each level  $\mathbf{F}_e^\ell$  are refined through attention-based skip connections, and then incorporated in the decoder path. The contribution of this work consists in the integration of a second-order geometric mechanism in the skip connections, leveraging similarities between features to learn more relevant and discriminative representations, thereby enhancing the attention of the segmentation network.

### 3.2 Geometric Riemannian learning

In this work, the proposed representation to summarize a bank of features is based on symmetric positive definite (SPD) matrices, capturing meaningful inter-channel relationships in a compact descriptor. To achieve this, given a feature tensor  $\mathbf{F} \in \mathbb{R}^{C \times H \times W \times D}$ , we first reshape  $\mathbf{F}$  into a rectangular matrix  $\mathbf{R} \in \mathbb{R}^{C \times N}$ , where  $N = H \times W \times D$ , such that each row of  $\mathbf{R}$  corresponds to a



single feature information. Then, the second-order descriptor is computed as:  $\mathbf{X}_0 = f_{SPDPool}(\mathbf{F}) = \frac{1}{N} \mathbf{R} \mathbf{R}^\top$ , resulting in a symmetric and positive definite (SPD) matrix  $\mathbf{X}_0 \in \mathbb{R}^{C \times C}$ . Here, each entry  $\mathbf{X}_0(i, j)$  encodes the correlation between the  $i$ -th and  $j$ -th features, capturing inter-channel relationships and summarizing relevant information from the input features.

Since SPD matrices lie on a Riemannian manifold, it is necessary to employ geometry-aware operations to preserve their structure and enable learning more meaningful representations from  $\mathbf{X}_0$ . Here, we utilized the *SPDnet*, preserving Riemannian manifold structure while learning lower-dimensional SPD discriminative representations [10]. For geometric learning, this network first projects SPD matrices into more compact SPD embeddings, following a *BiMap layer*, defined as:  $\mathbf{X}_k = f_{\text{BiMap}}(\mathbf{X}_{k-1}) = \mathbf{W}_k \mathbf{X}_{k-1} \mathbf{W}_k^\top$ , where  $\mathbf{X}_{k-1} \in \mathcal{S}_{++}^{d_{k-1}}$  is the SPD matrix from the previous layer, with dimension  $d_{k-1} \times d_{k-1}$ , and  $\mathbf{W}_k \in \mathbb{R}_*^{d_k \times d_{k-1}}$  is the transformation matrix that generates the new SPD matrix  $\mathbf{X}_k \in \mathcal{S}_{++}^{d_k}$ . Similar to conventional networks, the dimensions of the SPD matrices are progressively reduced ( $d_k < d_{k-1}$ ) via this bilinear mapping.

Resulting matrices  $\mathbf{X}_k$  may have eigenvalues near zero due to optimization and numerical procedures during the learning process. So, an eigenvalue rectification (*ReEig*) layer is considered [10]. It is defined as  $\mathbf{X}_k = f_{\text{ReEig}}(\mathbf{X}_{k-1}) = \mathbf{U}_{k-1} \max(\epsilon \mathbf{I}, \mathbf{\Sigma}_{k-1}) \mathbf{U}_{k-1}^\top$ , where  $\mathbf{U}_{k-1}$  and  $\mathbf{\Sigma}_{k-1}$  matrices correspond to the eigenvectors and eigenvalues, respectively,  $\mathbf{X}_{k-1} = \mathbf{U}_{k-1} \mathbf{\Sigma}_{k-1} \mathbf{U}_{k-1}^\top$ . Here,  $\epsilon \in \mathbb{R}$  denotes a non-negative rectification threshold. This operation prevents non-positive eigenvalues and preserves the SPD data structure [10]. The concatenation of a *BiMap* and a *ReEig* layer is referred to as a *BiRe* block. After a sequence of *BiRe* blocks, the geometric Riemannian descriptors are mapped back to an Euclidean space using the Riemannian logarithm map (*LogEig layer*):  $\mathbf{X}_k = f_{\text{Log}}(\mathbf{X}_{k-1}) = \mathbf{U}_{k-1} \log(\mathbf{\Sigma}_{k-1}) \mathbf{U}_{k-1}^\top$ . The resulting matrix  $\mathbf{X}_k$ , constitutes a lower-dimensional Euclidean descriptor of semantic inter-channel relationships. The resulting  $\mathbf{X}_k$  is a symmetric matrix, thereby its upper triangular part is used for further processing.

### 3.3 Second Order Geometric Attention module (SOGA)

Typically, attention mechanisms in U-Net like networks are implemented to guide encoder features ( $\mathbf{F}_e^\ell$ ), while using gating features from the decoder ( $\mathbf{F}_d^\ell$ ) for injecting semantic context and together help the network focus on regions relevant for the segmentation task. Specifically, the output of an attention mechanism is  $\hat{\mathbf{F}}^\ell = \mathbf{F}_e^\ell \odot \boldsymbol{\alpha}_\ell$ , where  $\boldsymbol{\alpha}_\ell$  is a learnable vector with attention coefficients that refine the encoder feature responses. These attention coefficients are typically computed from both encoder and decoder feature tensors,  $\mathbf{F}_e^\ell$  and  $\mathbf{F}_d^\ell$ , involving common fusion operations including concatenation, multiplication, or summation.

In this work, we propose including second-order information throughout the attention process. For this purpose, at each level  $\ell$  we first compute a SPD descriptor for both the encoder and decoder features banks as  $\mathbf{X}_* = f_{SPDPool}(\mathbf{F}_*^\ell)$ ,

with  $*$   $\in \{\mathbf{e}, \mathbf{d}\}$ . These descriptors, of dimension  $C_\ell \times C_\ell$ , capture relevant information from the encoder and decoder feature banks, encoding key semantic visual patterns propagated via skip connections that are related to PCa. From the SPD descriptors  $\mathbf{X}_\mathbf{e}$  and  $\mathbf{X}_\mathbf{d}$ , lower-dimensional geometric representations are obtained via two independent Riemannian SPD networks  $\varphi_\mathbf{e}$  and  $\varphi_\mathbf{d}$ . Each  $\varphi_*$  consists of a *BiRe* layer and a *LogEig* layer which outputs a symmetric matrix of dimension  $\frac{C_\ell}{2} \times \frac{C_\ell}{2}$ , and a final output layer that extracts the upper triangular part. The resulting encoder and decoder representations are concatenated to form a single geometric embedded representation:

$$\mathbf{e}_\ell = [\varphi_\mathbf{e}(f_{\text{SPDPool}}(\mathbf{F}_\mathbf{e}^\ell)), \varphi_\mathbf{d}(f_{\text{SPDPool}}(\mathbf{F}_\mathbf{d}^\ell))] . \quad (1)$$

These fused descriptors produce a vector of dimension  $C_\ell(\frac{C_\ell}{2} + 1)$ , aggregating discriminative information from the encoder and decoder via SPD descriptors and Riemannian processing. To obtain the attention coefficients, the embedding  $\mathbf{e}_\ell$  is passed through a nonlinear mapping composed of two fully connected layers:

$$\alpha_\ell = \delta(\mathbf{W}_\ell^2(W_\ell^1 \mathbf{e}_\ell + \mathbf{b}_\ell^1) + \mathbf{b}_\ell^2) . \quad (2)$$

Here,  $\mathbf{W}_\ell^1 \in \mathbb{R}^{\frac{C_\ell}{r} \times (C_\ell(\frac{C_\ell}{2} + 1))}$  and  $\mathbf{W}_\ell^2 \in \mathbb{R}^{C_\ell \times \frac{C_\ell}{r}}$  represent the linear transformations, and  $b_\ell^1 \in \mathbb{R}^{\frac{C_\ell}{r}}$ ,  $b_\ell^2 \in \mathbb{R}^{C_\ell}$  are the bias terms. The hyperparameter  $r$  controls the reduction ratio, aiming to balance representational capacity and parameter efficiency. We selected  $r = 4$ , as described in [9]. The output of this process is the vector of attention coefficients that recalibrates features as  $\hat{\mathbf{F}}^\ell = \mathbf{F}_\mathbf{e}^\ell \odot \alpha_\ell$ .

## 4 Dataset

Two different datasets were included in the validation of the proposed approach, among others, to assess the generalization properties of the geometrical introduced approach. Figure 2 describes the experimental protocol conducted to validate the approach from two considered datasets. Thereafter, we briefly describe both set of data.

*PI-CAI challenge dataset.* This dataset is the largest public benchmark for csPCa detection, comprising 1,500 public bp-MRI studies for public training and development. The organizers provided delineations of the prostate’s central and peripheral zones for each study. Of the 1500 cases, 1075 are benign and 425 are biopsy-confirmed csPCa. Among the csPCa cases, 220 include experts’ annotations of csPCa lesions, while the remaining 205 provided AI-generated delineations mentioned as “pseudo-labels”, which were used exclusively for training. To address imaging variability, the dataset includes bp-MRI scans from three Dutch centers: Radboud University Medical Center (RUMC), Ziekenhuis Groep Twente (ZGT), and University Medical Center Groningen [19].

The PI-CAI dataset includes an additional hidden cohort designed for models’ tuning. This cohort comprises 100 bp-MRI studies. The challenge organizers

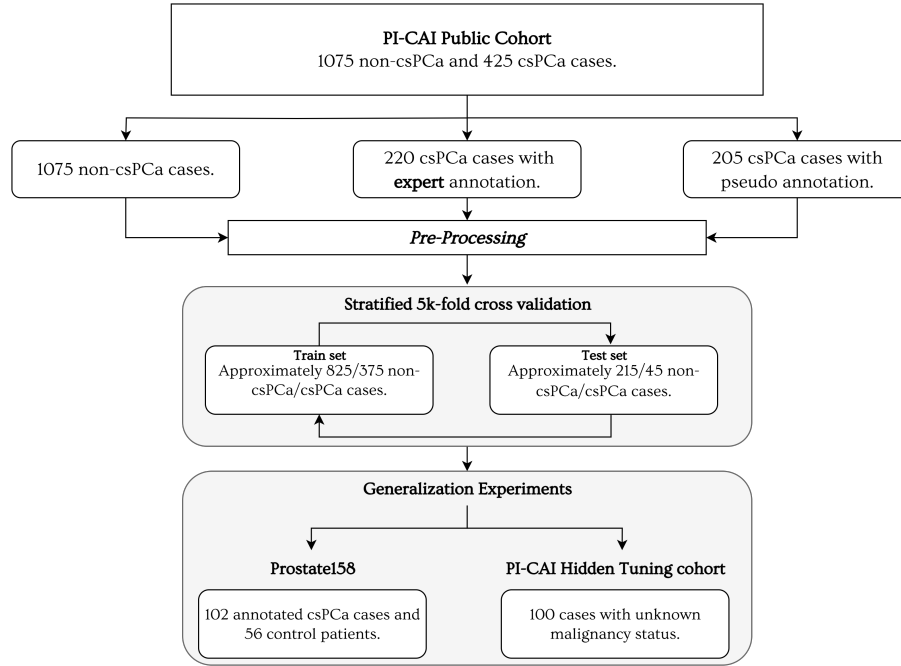


Fig. 2: Datasets flowchart. The training and validation of models was performed on the public PI-CAI cohort. For generalization evaluation we considered the hidden PI-CAI Tuning cohort and the Prostate158 dataset.

reserved the lesions' delineations for this set, and performance was assessed indirectly using metrics reported by the challenge evaluation platform. We used this cohort to report the generalization ability of models trained on the public training cohort (see Figure 2).

Pre-processing steps involved resampling the scans to a voxel resolution of  $0.5\text{mm} \times 0.5\text{mm} \times 3.0\text{mm}$ . The scans were then center-cropped to obtain 24 axial slices of  $384 \times 384$  pixels [19]. A volumetric region of interest (vROI) of size  $20 \times 180 \times 180$  voxels was extracted around the volumetric centroid of the prostate gland. Each input sequence, including T2-weighted images (T2W), diffusion-weighted images (DWI), and apparent diffusion coefficient maps (ADC), was independently normalized using z-score normalization.

*PROSTATE158.* This dataset comprises bp-MRI studies from 158 patients collected at Charité University Hospital Berlin [1]. Of these, 102 cases were biopsy- or surgery-confirmed csPCa, while the remaining 56 were cancer-free controls. The dataset is divided into 139 training and 19 testing cases, all annotated by two board-certified radiologists. Each case includes expert delineation of prostate zones (central and peripheral). The same pre-processing steps used for the PI-CAI public cohort was applied to this external cohort. We also used this indepen-

dent external cohort to evaluate the generalization of our models. The training and testing cases of PROSTATE158 were merged to form a single independent evaluation set. An overview of this workflow is provided in Figure 2.

## 5 Experimental Setup

We selected a standard 3D U-Net architecture to include the proposed *SOGA* attention module. The network consists of five hierarchical levels, with next channel dimensions: [32, 64, 128, 256, 512]. At each level, this architecture has a convolutional block with two 3D convolution layers (kernel size of  $3 \times 3 \times 3$ ), followed by batch normalization and a ReLU activation function. Downsampling was performed using 3D max pooling (see Figure 1). We refer to this model, without attention blocks, as *U-Net*. To compare the proposed Second-Order Geometric Attention (*SOGA*) we considered a baseline attention module, using Global Average Pooling. This method is referred as *U-Net FOA* because of the first order representation of features. Also, the proposed approach was compared with an architecture that included SPD pooling without the subsequent geometric processing, *i.e.*, without the incorporation of learnable BiRE blocks. This method is referred as *U-Net SOA* because of the second order description of features.

In the proposed *SOGA* method, a rectification threshold of  $\epsilon = 10^{-4}$  was applied in the ReEig layers. RMSprop with a learning rate of  $10^{-4}$  was used for non-Riemannian parameters, while BiMap weights were optimized via gradient descent on the Stiefel manifold with a learning rate of  $10^{-1}$  [10]. All models were trained for 200 epochs using the sum of Dice Loss and Binary Cross-Entropy as the loss function. In addition, we conducted experiments by integrating these methods into the nnU-Net framework [11]. Furthermore, transformer-based models: UNETR [8] and Swin UNETR [7] were adhered to the nnU-Net framework for comparison with state-of-the-art attention-based methods. In contrast to approaches relying solely on the original U-Net architecture, these nnU-Net-based methods incorporate automated self-configuring capabilities, such as dynamic data adaptation, offering a more robust and generalizable training pipeline [11].

*Evaluation.* To evaluate the performance of studied methods in segmentation and detection of csPCa, multiple metrics were considered, providing patient-level and lesion-level performance. Before computing metrics, detection maps and confidence scores (maximum prediction values from these maps) were obtained using PICA’s recommended post-processing [19]. We considered the following evaluation metrics: the AUC-ROC, the Average Precision (AP), a standard metric in detection tasks that quantifies the trade-off between precision and recall across confidence thresholds. Precision decreases as false positives increase, so a higher AP reflects better performance. Detections were considered correct with an IoU of at least 0.1, consistent with literature [12]. A critical issue in medical scenarios is the oversight of lesions. To address this, we included the Sensitivity at 1 False Positive (SEN@1FP) score. This metric quantifies the sensitivity at a fixed tolerance of one false positive per case, thereby reflecting the impact of

false negatives, as lower sensitivity indicates a higher number of missed lesions. The Dice Similarity Coefficient (DSC) was used for measuring the segmentation performance.

In the PI-CAI challenge experiments, we followed the official 5-fold cross-validation using the predefined splits provided by the organizers [21]. Ensemble predictions for the PI-CAI Hidden Tuning and Prostate158 cohorts were obtained by averaging the softmax outputs across folds. To assess statistical significance, we employed the Delong test to compare the AUCs of the different models. This non-parametric test evaluates the null hypothesis that their AUCs are not significantly different by estimating the covariance of correlated ROC curves.

## 6 Results

### 6.1 Detection in PI-CAI Public Cohort

Results in the PI-CAI Public Cohort is reported in Table 1. Interestingly, the integration of the *SOA* module (*U-Net SOA*) achieved the highest overall performance, while the proposed *SOGA* module (*U-Net SOGA*) demonstrated consistent improvements across all evaluation metrics when compared to both the baseline U-Net and its FOA-enhanced counterpart. This integration of the *SOGA* module increased the *U-Net* performance in AUC-ROC from 0.76 to 0.82 which was statistically significant ( $p < 0.05$ ), indicating a better capability of the model to distinguish between csPCa and non-csPCa cases. Besides, a wide improvement was observed analyzing the Sen@1FP score, increasing from 0.61 to 0.74 (+21.3%). Additionally, the DSC increased by 0.08 (+23.5%) and showed a reduction in its standard deviation.

As expected, using nnU-Net improved the performance of U-Net based methods, demonstrating the advantage of this framework for adjusting the data given its variability and configuring the training pipelines for the detection and segmentation task. Even, when *SOGA* module is integrated into the nnU-Net, it demonstrated it improves the detection performance, with an AP score of  $0.37 \pm 0.05$  (+5.7%), and in segmentation, with a DSC score of  $0.52 \pm 0.02$  (+23.8%). The AUC-ROC remained similar to the baseline (no statistical difference,  $p = 0.695$ ). Contrarily, the transformer-based models UNETR and Swin UNETR achieved lower overall performance than the proposed *SOGA*, with no statistical significance ( $p = 0.140$  and  $p = 0.067$ , respectively), and only slightly improve the performance of the baseline nnU-Net approach in AP and DSC scores.

### 6.2 Generalization evaluation

Regarding **PI-CAI Hidden Tuning cohort**, the challenge only provided AUC-ROC and AP scores. For this experiment, we used the nnU-Net framework as the *baseline* and evaluate performance when attention modules are incorporated,

Table 1: **Performance on PI-CAI public cohort.** Comparison between the proposed *SOGA* module (marked with \*) and attention-based models in a U-Net configuration (top) and using the nnU-Net framework (bottom). The best models from both configurations have their performance highlighted in **bold**.

| Model                    | AUC-ROC $\uparrow$                | AP $\uparrow$                     | Sen@1FP $\uparrow$                | DSC $\uparrow$                    |
|--------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| U-Net                    | $0.76 \pm 0.03$                   | $0.31 \pm 0.05$                   | $0.61 \pm 0.14$                   | $0.34 \pm 0.09$                   |
| U-Net FOA                | $0.70 \pm 0.04$                   | $0.21 \pm 0.08$                   | $0.49 \pm 0.11$                   | $0.25 \pm 0.08$                   |
| U-Net SOA                | <b><math>0.85 \pm 0.02</math></b> | <b><math>0.38 \pm 0.05</math></b> | <b><math>0.73 \pm 0.07</math></b> | $0.22 \pm 0.26$                   |
| U-Net SOGA*              | $0.82 \pm 0.03$                   | $0.35 \pm 0.08$                   | $0.74 \pm 0.06$                   | $0.42 \pm 0.04$                   |
| <i>nnU-Net framework</i> |                                   |                                   |                                   |                                   |
| Baseline                 | <b><math>0.83 \pm 0.04</math></b> | $0.33 \pm 0.09$                   | $0.82 \pm 0.04$                   | $0.47 \pm 0.03$                   |
| UNETR                    | $0.82 \pm 0.03$                   | $0.35 \pm 0.08$                   | $0.82 \pm 0.02$                   | $0.49 \pm 0.03$                   |
| Swin UNETR               | $0.80 \pm 0.03$                   | $0.34 \pm 0.07$                   | <b><math>0.83 \pm 0.04</math></b> | $0.50 \pm 0.03$                   |
| FOA                      | <b><math>0.83 \pm 0.03</math></b> | $0.36 \pm 0.70$                   | $0.81 \pm 0.05$                   | $0.51 \pm 0.04$                   |
| SOA                      | $0.82 \pm 0.04$                   | $0.35 \pm 0.10$                   | <b><math>0.83 \pm 0.07</math></b> | $0.51 \pm 0.02$                   |
| SOGA*                    | <b><math>0.83 \pm 0.03</math></b> | <b><math>0.37 \pm 0.05</math></b> | $0.82 \pm 0.02$                   | <b><math>0.52 \pm 0.02</math></b> |

as shown in Table 2. Here, AUC-ROC scores indicate a similar classification performance across all methods. Nevertheless, the incorporation of the proposed *SOGA* attention module led to a notable improvement in the AP score, increasing from 0.37 to 0.46 (+24.3%). This difference in detection performance is higher than observed in the PI-CAI public cohort when *SOGA* was incorporated in nnU-Net, indicating that the proposed *SOGA* module robust on unseen cases. This substantial enhancement highlights the method’s generalization ability in unseen cases. Moreover, Swin UNETR stood out as the top-performing approach, achieving an AP of 0.54, representing an improvement of 46% compared to the baseline.

Table 2: **Performance on PI-CAI Hidden Tuning cohort.** Comparison of the *baseline* nnU-Net, nnU-Net with *FOA* attention, nnU-Net with proposed *SOGA* attention module, and Swin UNETR.

| Model      | AUC-ROC $\uparrow$ | AP $\uparrow$ |
|------------|--------------------|---------------|
| Baseline   | <b>0.78</b>        | 0.37          |
| UNETR      | 0.75               | 0.47          |
| Swin UNETR | <b>0.78</b>        | <b>0.54</b>   |
| FOA        | 0.76               | 0.40          |
| SOA        | 0.71               | 0.44          |
| SOGA*      | 0.77               | 0.46          |

To further assess the generalization performance, we conducted additional evaluations on an independent external cohort, **Prostate158** dataset. Results are shown in Table 3. Here, attention-based models consistently demonstrated

superior generalization performance, followed by UNETR and Swin UNETR methods. The Swin UNETR method had a significantly different AUC-ROC compared to the proposed method ( $p < 0.05$ ). However, the difference for the UNETR method was not significant ( $p = 0.140$ ). In comparison with the baseline nnU-Net, including *SOGA* leads to an improvement of 0.13 points (+21.0%) in AUC-ROC, which is statistically significant ( $p < 0.05$ ), 0.22 (+146.6%) in AP, 0.29 (+126.1%) in Sen@1FP, and 0.19 (+158.3%) in DSC. These results underscore the ability of *SOGA* module to support detection and segmentation of csPCa lesions, more remarkably in cases from external and independent cohorts. This can be explained by the ability of the proposed method to more effectively capture visual features relevant to identifying csPCa lesions, which allows for better performance in novel external, and independent datasets.

Table 3: **Performance on the Prostate158 dataset.** Comparison of the *baseline* nnU-Net and attention-based methods adhered into this framework.

| Model        | AUC-ROC $\uparrow$ | AP $\uparrow$ | Sen@1FP $\uparrow$ | DSC $\uparrow$ |
|--------------|--------------------|---------------|--------------------|----------------|
| Baseline     | 0.62               | 0.15          | 0.23               | 0.12           |
| UNETR        | 0.73               | 0.29          | 0.38               | 0.23           |
| Swin UNETR   | 0.70               | 0.27          | 0.38               | 0.23           |
| FOA          | 0.67               | 0.18          | 0.31               | 0.18           |
| SOA          | 0.72               | 0.29          | 0.43               | 0.24           |
| <b>SOGA*</b> | <b>0.75</b>        | <b>0.37</b>   | <b>0.52</b>        | <b>0.31</b>    |

Additionally, we calculated the models’ performance differentiating by lesions’ size. To this end, lesions in the *Prostate158* dataset were stratified by lesion volume as: *small* ( $< 931 \text{ mm}^3$ ), *medium* ( $931\text{--}2337 \text{ mm}^3$ ), and *large* ( $>2337 \text{ mm}^3$ ), using the 33rd and 66th percentiles as thresholds. As shown in Table 4, the proposed *SOGA* module consistently outperformed baseline nnU-Net and alternative attention-based methods across all lesion sizes. For small lesions, typically the most challenging to detect, *SOGA* achieved the best performance, followed in order by SOA, UNETR, Swin UNETR, FOA, and baseline nnU-Net. In the group of medium-sized lesions, a similar order of performance is observed. Additionally, improved detection and segmentation scores are reported for all models when compared to their corresponding scores in the small lesion group. For large lesions, all methods achieved higher performance compared to medium and small lesions. While *SOGA* continued to demonstrate strong overall performance, it was surpassed in AP and Sen1FP scores by Swin UNETR. Notably, the advantage of the proposed *SOGA* was more pronounced for small and medium-sized lesions.

Figure 3 presents a visual comparison between expert annotations and models’ predictions for three random cases. In case a), there is a large lesion, delineated only by Swin UNETR, SOA and the proposed method *SOGA*, with a slight advantage in segmentation by *SOGA* of 3 points in the DSC, and a

Table 4: **Performance across different lesion sizes.** Comparison between the *baseline* nnU-Net and attention-based methods for the detection and segmentation of small, medium, and large lesions in the *Prostate158* dataset.

| Metric        | Baseline | UNETR | S. UNETR    | FOA  | SOA  | SOGA*       |
|---------------|----------|-------|-------------|------|------|-------------|
| <i>Small</i>  |          |       |             |      |      |             |
| DSC           | 0.10     | 0.14  | 0.11        | 0.11 | 0.18 | <b>0.22</b> |
| AP            | 0.04     | 0.15  | 0.12        | 0.08 | 0.25 | <b>0.38</b> |
| Sen@1FP       | 0.14     | 0.28  | 0.20        | 0.17 | 0.31 | <b>0.50</b> |
| <i>Medium</i> |          |       |             |      |      |             |
| DSC           | 0.17     | 0.28  | 0.28        | 0.24 | 0.30 | <b>0.33</b> |
| AP            | 0.26     | 0.38  | 0.36        | 0.33 | 0.42 | <b>0.44</b> |
| Sen@1FP       | 0.28     | 0.41  | 0.41        | 0.37 | 0.48 | <b>0.50</b> |
| <i>Large</i>  |          |       |             |      |      |             |
| DSC           | 0.16     | 0.26  | 0.31        | 0.24 | 0.29 | <b>0.40</b> |
| AP            | 0.24     | 0.37  | <b>0.47</b> | 0.27 | 0.35 | 0.43        |
| Sen@1FP       | 0.25     | 0.43  | 0.50        | 0.36 | 0.48 | <b>0.57</b> |

medium-sized lesion that was missed by all models in all predicted 3D volumes. The segmentation is not visible in the current slice, nor in the adjacent ones. Upon closer inspection, it seems the lesion appears split into two separate regions in this particular slice, which may have contributed to the models failing to detect it as a single, unified structure. Case b) presents a medium-sized lesion detected by all models. The baseline nnU-Net only captured a single portion of the lesion, in contrast with the attention-based methods. Notably, *SOGA* shows an advantage to detect lesion and also delineate better its morphology, obtaining a higher DSC. Case c) presents a small lesion in the peripheral zone. This lesion was weakly detected by the nnU-Net and SOA models, and missed entirely by the FOA model, but was successfully identified by the UNETR, Swin UNETR, and *SOGA* methods. The *SOGA* method provided a more precise delineation and confident detection, demonstrating the advantage of employing second-order attention over first-order statistics in proposed attention module.



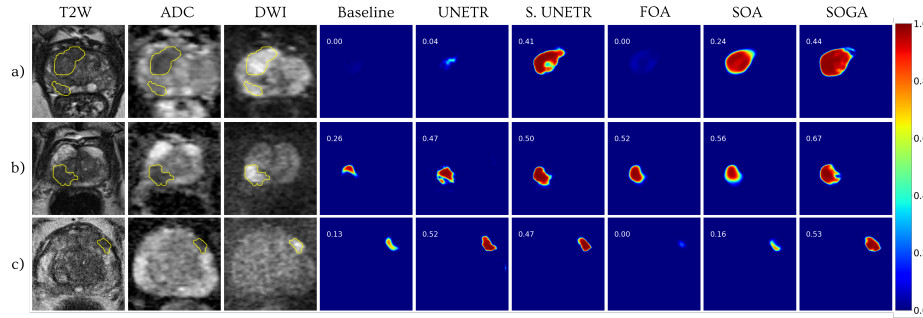


Fig 3: **Visualization of predictions** for three cases from the Prostate158 dataset, each presented in one row. The first three columns show bp-MRI images with ground-truth annotations. The predictions include DSC values.

## 7 Discussion and conclusions remarks

This study proposed a second-order geometric attention (*SOGA*) module to capture and enhance relevant information into encoder-decoder skip connections, enhancing the segmentation and detection of PCa lesions in bp-MRI. The proposed geometric attention leverages pairwise feature relationships from Symmetric Positive Definite (SPD) descriptors within a Riemannian manifold. Extensive evaluations demonstrated the advantage of including the proposed attention in standard U-Net and nnU-Net architectures. This indicated that the second-order descriptors along with the geometric processing utilized in the *SOGA* module effectively captured relevant patterns of PCa and improved the robustness of the baseline detection and segmentation architectures. The proposed *SOGA* report a remarked computational cost, during training, associated with operations on Riemannian manifolds, including eigenvalue decomposition and geometry-aware bilinear mapping.

The integration of second-order attention mechanisms over the nnU-Net led to performance generalization improvements over baseline architectures, further improving the performance on external data, obtaining an AUC-ROC of 0.83, Sen@1FP of 0.74, and a DSC of 0.42. Notably, the proposed method's advantage over the baseline and attention-based approaches was even more pronounced on an entirely independent cohort, surpassing by 21.0% in AUC-ROC, 126.1% in Sen@1FP, and 158.3% in DSC, the baseline nnU-Net model. This demonstrates its ability to capture relevant, robust features of prostate cancer and maintain a robust performance in new scenarios. This is significant for future real clinical adoption in new samples and at different clinical centers. Besides, the proposed approach offers a substantial contribution to the detection and segmentation of small lesions, key on early detection, which constitutes the most difficult challenges for both expert radiologists and deep learning models.

Future work should explore integration in novel detection frameworks, studying the relevance of geometric processing within attention mechanisms. In line

with generalization results, it is also important to evaluate the proposed module's capability in data-constrained scenarios and conduct additional studies expanding the number of datasets.

## References

1. Adams, L.C., Makowski, M.R., Engel, G., Rattunde, M., Busch, F., Asbach, P., Niehues, S.M., Vinayahalingam, S., van Ginneken, B., Litjens, G., et al.: Prostate158-an expert-annotated 3t mri dataset and algorithm for prostate cancer detection. *Computers in Biology and Medicine* **148**, 105817 (2022)
2. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**(3), 229–263 (2024)
3. Debs, N., Routier, A., Bône, A., Rohé, M.M.: Evaluation of a deep learning prostate cancer detection system on biparametric mri against radiological reading. *European Radiology* pp. 1–10 (2024)
4. Duran, A., Dussert, G., Rouvière, O., Jaouen, T., Jodoin, P.M., Lartizien, C.: Prostatattention-net: A deep attention model for prostate cancer segmentation by aggressiveness in mri scans. *Medical Image Analysis* **77**, 102347 (2022)
5. Gatti, M., Faletti, R., Callaris, G., Giglio, J., Berzovini, C., Gentile, F., Marra, G., Misischi, F., Molinaro, L., Bergamasco, L., et al.: Prostate cancer detection with biparametric magnetic resonance imaging (bpmri) by readers with different experience: performance and comparison with multiparametric (mpmri). *Abdominal Radiology* **44**, 1883–1893 (2019)
6. Hafiz, A.M., Parah, S.A., Bhat, R.U.A.: Attention mechanisms and deep learning for machine vision: A survey of the state of the art. *arXiv preprint arXiv:2106.07550* (2021)
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: *International MICCAI brainlesion workshop*. pp. 272–284. Springer (2021)
8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 574–584 (2022)
9. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7132–7141 (2018)
10. Huang, Z., Van Gool, L.: A riemannian network for spd matrix learning. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 31 (2017)
11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
12. Karagoz, A., Alis, D., Seker, M.E., Zeybel, G., Yergin, M., Oksuz, I., Karaarslan, E.: Anatomically guided self-adapting deep neural network for clinically significant prostate cancer detection on bi-parametric mri: a multi-center study. *Insights into Imaging* **14**(1), 110 (2023)
13. Li, W., Zheng, B., Shen, Q., Shi, X., Luo, K., Yao, Y., Li, X., Lv, S., Tao, J., Wei, Q.: Adaptive window adjustment with boundary dou loss for cascade segmentation of anatomy and lesions in prostate cancer using bpmri. *Neural Networks* **181**, 106831 (2025)

14. Li, Y., Wynne, J., Wang, J., Qiu, R.L., Roper, J., Pan, S., Jani, A.B., Liu, T., Patel, P.R., Mao, H., et al.: Cross-shaped windows transformer with self-supervised pretraining for clinically significant prostate cancer detection in bi-parametric mri. *Medical Physics* **52**(2), 993–1004 (2025)
15. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60–88 (2017)
16. Naji, L., Randhawa, H., Sohani, Z., Dennis, B., Lautenbach, D., Kavanagh, O., Bawor, M., Banfield, L., Profetto, J.: 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)
17. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18. pp. 234–241. Springer (2015)
19. Saha, A., Bosma, J.S., Twilt, J.J., van Ginneken, B., Bjartell, A., Padhani, A.R., Bonekamp, D., Villeirs, G., Salomon, G., Giannarini, G., et al.: Artificial intelligence and radiologists in prostate cancer detection on mri (pi-cai): an international, paired, non-inferiority, confirmatory study. *The Lancet Oncology* **25**(7), 879–887 (2024)
20. Saha, A., Hosseinzadeh, M., Huisman, H.: End-to-end prostate cancer detection in bpmri via 3d cnns: effects of attention mechanisms, clinical priori and decoupled false positive reduction. *Medical image analysis* **73**, 102155 (2021)
21. Saha, A., Twilt, J.J., Bosma, J.S., van Ginneken, B., Yakar, D., Elschot, M., Veltman, J., Fütterer, J., de Rooij, M., Huisman, H.: The pi-cai challenge: public training and development dataset. *Zenodo*, Jun (2022)
22. Shoag, J., Barbieri, C.E.: Clinical variability and molecular heterogeneity in prostate cancer. *Asian journal of andrology* **18**(4), 543–548 (2016)
23. Thompson, J., Lawrentschuk, N., Frydenberg, M., Thompson, L., Stricker, P., Usanz: The role of magnetic resonance imaging in the diagnosis and management of prostate cancer. *BJU international* **112**, 6–20 (2013)
24. Wei, C., Liu, Z., Zhang, Y., Fan, L.: Enhancing prostate cancer segmentation in bpmri: Integrating zonal awareness into attention-guided u-net. *Digital Health* **11**, 20552076251314546 (2025)
25. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*. pp. 3–19 (2018)
26. Yang, C., Li, B., Luan, Y., Wang, S., Bian, Y., Zhang, J., Wang, Z., Liu, B., Chen, X., Hacker, M., et al.: Deep learning model for the detection of prostate cancer and classification of clinically significant disease using multiparametric mri in comparison to pi-rads score. In: *Urologic Oncology: Seminars and Original Investigations*. vol. 42, pp. 158–e17. Elsevier (2024)