MDPI

Article

Neural Network Models for Prostate Zones Segmentation in Magnetic Resonance Imaging

Saman Fouladi [1,*,†], Luca Di Palma [2,†], Fatemeh Darvizeh [2], Deborah Fazzini [2], Alessandro Maiocchi [3], Sergio Papa [2], Gabriele Gianini [4,*] and Marco Alì [2,3,*]

1 Department of Computer Science, Università degli Studi di Milano, 20133 Milano, Italy
2 Department of Diagnostic Imaging and Stereotactic Radiosurgery, Centro Diagnostico Italiano (CDI), 20147 Milano, Italy; luca.dipalma@cdi.it (L.D.P.); fatemeh.darvizeh@cdi.it (F.D.); deborah.fazzini@cdi.it (D.F.); sergio.papa@cdi.it (S.P.)
3 R&D, Bracco Imaging S.p.A., 20134 Milano, Italy; alessandro.maiocchi@bracco.com
4 Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), Università degli Studi di Milano-Bicocca, 20126 Milano, Italy
* Correspondence: saman.fouladi@unimi.it (S.F.); gabriele.gianini@unimib.it (G.G.); marco.ali@cdi.it (M.A.)
† These authors contributed equally to this work.

Abstract: Prostate cancer (PCa) is one of the most common tumors diagnosed in men worldwide, with approximately 1.7 million new cases expected by 2030. Most cancerous lesions in PCa are located in the peripheral zone (PZ); therefore, accurate identification of the location of the lesion is essential for effective diagnosis and treatment. Zonal segmentation in magnetic resonance imaging (MRI) scans is critical and plays a key role in pinpointing cancerous regions and treatment strategies. In this work, we report on the development of three advanced neural network-based models: one based on ensemble learning, one on Meta-Net, and one on YOLO-V8. They were tailored for the segmentation of the central gland (CG) and PZ using a small dataset of 90 MRI scans for training, 25 MRIs for validation, and 24 scans for testing. The ensemble learning method, combining U-Net-based models (Attention-Res-U-Net, Vanilla-Net, and V-Net), achieved an IoU of 79.3% and DSC of 88.4% for CG and an IoU of 54.5% and DSC of 70.5% for PZ on the test set. Meta-Net, used for the first time in segmentation, demonstrated an IoU of 78% and DSC of 88% for CG, while YOLO-V8 outperformed both models with an IoU of 80% and DSC of 89% for CG and an IoU of 58% and DSC of 73% for PZ.

Keywords: prostate zonal segmentation; ensemble learning; Meta-Net; YOLO-V8; U-Net neural network; magnetic resonance imaging (MRI)

1. Introduction

Prostate cancer (PCa) is one of the most common types of tumors diagnosed in men worldwide [1]. By 2030, it is predicted that approximately 1.7 million new cases will be diagnosed globally [2]. PCa originates within the prostate gland, which is categorized into various functional zones: the central zone (CZ), the peripheral zone (PZ), and the transitional zone (TZ) [3]. The PZ, which extends posterolaterally from the apex to the base of the gland, is particularly significant as it is the area most frequently affected by prostate carcinomas. This prevalence is due to the PZ containing the majority of the prostatic glandular tissue [4,5]. Prostate cancers that develop in the PZ account for over 70% of all prostate cancer cases and are associated with worse clinical outcomes compared to those arising in the TZ [6]. Therefore, precise segmentation of these zones, particularly the PZ, in

magnetic resonance imaging (MRI) scans is vital for the effective diagnosis and treatment of prostate cancer [7].

MRI is the primary imaging method for identifying and localizing prostate cancer, as outlined by the Prostate Imaging Reporting and Data System (PI-RADS) scoring framework. This approach relies on an understanding of the zonal anatomy of the prostate, which is essential for accurate cancer detection. The PI-RADS scores vary depending on the regions assessed; diffusion-weighted imaging (DWI) is utilized for lesions in the PZ, while T2-weighted (T2W) imaging is employed for transitional zones. Additionally, zonal segmentation is vital for several clinical applications, including consistent evaluation of prostate volume and prostate-specific antigen (PSA) density, MRI–ultrasound fusion biopsy, radiotherapy, and focal treatment planning [7].

Prostate zonal segmentation is conventionally carried out manually on T2W images by delineating the prostate slice by slice. This method is highly time-consuming and laborious, often resulting in significant inter- and intra-observer variability. The variability arises from the subjective nature of human interpretation of organ boundaries and the considerable differences in prostate anatomy and gland intensity heterogeneity among patients [8]. There is a critical demand for the development of automated techniques to streamline the prostate segmentation process, ensuring both speed and precision. Moreover, the identification and staging of prostate cancer on MRI are dependent on precise segmentation of the prostate zones [9,10].

Automating the zonal segmentation of the prostate presents significant challenges for several reasons. The prostate gland exhibits considerable morphological variation, intra-prostatic heterogeneity, and often poor contrast with surrounding tissues, complicating the delineation of its zonal contours. Additionally, assessing the applicability of these methods across multiple institutions can be difficult due to substantial technical variability in image acquisition. Factors such as inconsistent MRI signal intensity, differences in acquisition protocols, field strength, scanner types, and coil configurations all significantly influence image characteristics [11].

In this article, we introduce three innovative neural network approaches—ensemble learning, MetaNet, and YOLO-V8—for effective segmentation of the CG and PZ regions in T2-weighted MRI scans. The ensemble method leverages the combined strengths of three U-Net-based models, Attention-Res-U-Net (Att-R-Net), Vanilla-Net, and V-Net, by averaging their outputs. Notably, our work marks the first implementation of MetaNet, specifically adapted for segmentation tasks. Additionally, YOLO-V8 demonstrates strong performance, showcasing significant improvements in its architecture to optimize segmentation accuracy. This paper aims to advance the state of the art in medical image segmentation.

2. Related Work

This retrospective study [12] compared deep learning methods for prostate segmentation using 204 patients from the PROSTATEx dataset with 3T T2-weighted images. Manual segmentations of different prostate zones by four operators served as the basis for training the U-net, ENet, and ERFNet models. ENet achieved the highest accuracy, with Dice similarity coefficient scores of 91% for the whole gland, 87% for the TZ, and 71% for the PZ. U-net and ERFNet showed slightly lower performance. ENet also had the lowest training and inference times, demonstrating that deep learning can effectively segment the prostate on T2-weighted images.

In [13], a BASC-Net is designed to automatically segment prostate zones from MRI, enhancing prostate cancer diagnosis by accurately delineating the PZ and central gland (CG). Its architecture includes a semantic clustering attention (SCA) module for feature extraction and attention map creation, along with a boundary-aware contrastive (BAC)

loss that improves feature similarity within the same zone while distinguishing between different zones. In evaluations on the NCI-ISBI 2013 Challenge and Prostate158 datasets, BASC-Net outperformed nine competing methods, achieving Dice similarity coefficients of 79.9% for PZ and 88.6% for CG on the NCI-ISBI dataset and 80.5% for PZ and 89.2% for CG on Prostate158. These results demonstrate BASC-Net's potential to improve prostate lesion detection through more accurate zonal segmentation.

The authors in [14] discuss the creation and assessment of ensemble models aimed at segmenting prostate zones (the anatomical model) and identifying pathologically suspicious areas (the detection model). A key innovation of this approach is the integration of pre-training within the standard nnU-Net framework, coupled with a revised loss function that accounts for the variability in expert annotations. The anatomical segmentation model demonstrated impressive performance, achieving a Dice score (DSC) of 0.915 for the prostate gland, 0.865 for the transition zone, and 0.736 for the PZ.

In [15], a 3D U-Net model was developed and tested on data from 223 patients, including an internal group of 93 and two external datasets (ETDpub, $n = 141$ and ETDpri, $n = 59$). The model's performance was evaluated using DSCs, 95th Hausdorff distance (95HD), and average boundary distance (ABD), and it was compared to a junior radiologist's results. The DSCs were 0.909, 0.889, and 0.869 for the CG and 0.844, 0.755, and 0.764 for the transition zone (PZ) across the datasets. The model outperformed the radiologist in PZ segmentation (DSC of 0.769 vs. 0.706) and volume estimation. Important factors influencing performance included CG volume and MR vendor. Overall, the 3D U-Net model demonstrated effective auto-segmentation for prostate anatomy.

A new model called convolution coupled Transformer U-Net (CCT-Unet) has been developed for prostate segmentation, combining the strengths of Transformer-based models and convolutional neural networks (CNNs). While Transformers excel at capturing the global context, they often struggle with small prostate MRI datasets due to local variation insensitivity. CCT-Unet integrates a convolutional embedding block to maintain edges and a convolution Transformer block for better local feature extraction. Testing on the ProstateX and Huashan datasets showed that CCT-Unet outperformed existing methods, with Dice coefficient scores of 80.39% for the PZ and 87.49% for the transition zone [16].

In [17], a heterogeneous dataset comprising 243 T2-weighted prostate MRI studies from seven countries, utilizing ten machines from three different vendors, was employed to train and test a U-Net-based model with deep supervision and cyclical learning rate adjustments. Two experienced radiologists manually delineated the central gland-transition zone (CZ-TZ), PZ, and seminal vesicles (SVs) as ground truth. The model's performance was assessed using the DSC, with scores above 0.7 deemed accurate. On testing with 120 studies, the model achieved DSC values of 0.88 ± 0.01 for the prostate gland, 0.85 ± 0.02 for the CZ-TZ, and 0.72 ± 0.02 for both the PZ and SV.

In [18], an automated machine learning model was developed to segment the prostate gland, PZ, and transition zone (TZ) using MRI. This study involved consecutive men undergoing prostate MRI and biopsy, with images manually segmented by experienced radiologists. A novel two-stage Green Learning (GL) model was designed, where the first stage segments the prostate gland, and the second stage delineates the TZ and PZ. The project included 119 patients and 19,992 T2-weighted images, with a training dataset of 95 MRIs. The model achieved mean Dice scores of 0.85 for the whole prostate, 0.62 for the PZ, and 0.81 for the TZ, along with Pearson correlation coefficients of 0.92, 0.62, and 0.93 for volume accuracy, respectively (all $p < 0.01$). This work demonstrates the effectiveness of the ML model in automated prostate segmentation and includes a user-friendly web interface for annotation adjustments.

Ref. [19] introduced a 2D–3D convolutional neural network ensemble, PPZ-SegNet, designed to automatically segment the prostate gland and PZ using T2-weighted MRI sequences. It utilized four public datasets, including Train 1 and Test 1, with the latter derived from the same cohort, alongside Test 2, Test 3, and Test 4. The anatomical structures were manually delineated by a radiologist, except for Test 4, which utilized pre-marked anatomy. The model, constructed through Bayesian hyperparameter optimization and trained on 150 cases, was evaluated on an independent set of 283 T2W MRI prostate cases without further tuning. Sourced from the Cancer Imaging Archive (TCIA), segmentation performance was measured using the Dice similarity coefficient and Hausdorff distance, with average Dice scores of 0.86, 0.79, 0.81, and 0.62 across the test sets.

A new neural network called Dense U-net was developed for automatic segmentation of the prostate and its zones, blending concepts from DenseNet and U-net. It was trained on 141 patient datasets and tested on 47, utilizing axial T2-weighted images. The network demonstrated effective segmentation capabilities, even with imprecise labels. Compared to U-net, Dense-2 U-net achieved higher average Dice scores: 92.1% for the whole prostate, improving upon U-net's 90.7%, and 78.1% for the PZ [20].

Study [21] introduced a method for segmenting the whole prostate gland (WG), PZ, and CG using an apparent diffusion coefficient (ADC) and T2-weighted images. The approach employed two models with two U-Nets each, trained on a dataset of 225 patients. The results showed high performance on the test dataset, achieving Dice similarity coefficients of 95.33 for WG, 93.75 for CG, and 86.78 for PZ with T2W images. For ADC images, the DSC values were 92.09 for WG, 89.89 for CG, and 86.1 for PZ. Table 1 provides an overview of the results and findings from previous studies on prostate zonal segmentation.

Table 1. The summary of the previous studies for prostate zonal segmentation.

Study	Network	Dataset	CG		PZ	
[12]	ENet	PROSTATEx (204 MRIs)	DSC:	87%	DSC:	71%
[13]	BASC-Net	NCI-ISBI 2013 (80 MRIs)	DSC:	88.6%	DSC:	79.9%
		Prostate158 (102 MRIs)	DSC:	89.2%	DSC:	80.5%
[14]	nnU-Net	Private dataset (607 MRIs)	DSC:	86.5%	DSC:	73.6%
[15]	3D U-Net	From 223 patients, including an internal group of 93, and two external datasets (ETDpub, n = 141 and ETDpri, n = 59)	DSC:	86.9%	DSC:	76.9%
[16]	CCT-Unet	ProstateX and Huashan datasets (240 MRIs)	DSC:	87.49%	DSC:	80.39%
[17]	U-Net-based model	Inhouse dataset containing 243 T2W	DSC:	85%	DSC:	72%
[18]	A novel two-stage Green Learning	A dataset containing 119 MRIs	DSC:	81%	DSC:	62%
[19]	A 2D–3D convolutional neural network ensemble (PPZ-SegNet)	Cancer Imaging Archive (training: 150 and Test: 283 MRIs)	Not reported		DSC:	62%
[20]	Dense U-net	A dataset containing 141 MRIs	Not reported		DSC:	78.1%
[21]	U-Net	A dataset containing 225 MRIs (T2W)	DSC:	93.75%	DSC:	86.78%
		A dataset containing 225 MRIs (ADC)	DSC:	89.89%	DSC:	86.1%

3. Materials and Methods

3.1. Data Acquisition

For this study, the Prostate158 dataset was utilized. The Prostate158 dataset includes 158 annotated multi-parametric 3T prostate MRIs, featuring T2W and diffusion-weighted (DW) sequences, along with extracted ADC maps. These MRIs were obtained using a Siemens VIDA and Skyra 3T clinical scanner, adhering to established guidelines and protocols that included B1 calibration. The T2W sequences had a slice thickness of 3 mm, no inter-slice gap, and an in-plane resolution of 0.47×0.47 mm. The DWI was similarly captured with a slice thickness of 3 mm and an in-plane resolution of 1.4×1.4 mm. ADC maps were generated using b-values ranging from 50 to 1000 s/mm^3, with some up to 1400 s/mm^3, utilizing pre-installed scanner software (version VE11A). Following image acquisition and anonymization, the T2W images and ADC maps were stored on a local PACS and then moved to an internal server. Segmentation was performed with the open-source software ITK-Snap (version 3.8.0) by two experienced, board-certified radiologists, who provided detailed pixel-by-pixel annotations for the CG, PZ, and prostate cancer (PCa) lesions using the axial T2W sequence [22]. Out of the 158 MRIs, 139 T2W sequences included zonal masks, and these were the ones we utilized.

3.2. Neural Network Architecture and Training

In this study, we implemented three different approaches for prostate zonal segmentation. The first approach involved using 5-fold cross-validation to train three separate models: Att-R-Net, Vanilla-Net, and V-Net. As a result of the training process, we developed a total of 15 distinct models. Subsequently, we employed average ensemble learning across these 15 models to enhance prostate zonal segmentation accuracy.

The second approach involved leveraging the innovative MetaNet V1 [23] architecture, which had not previously been applied to segmentation tasks. This marks the first instance of using MetaNet (version 1) for prostate zonal segmentation. In this approach, we incorporated models from our ensemble learning method and integrated them through MetaNet. Afterward, we identified the top-performing combination of MetaNet models based on their performance on the validation set, which we subsequently tested on the test set. MetaNet's capacity to enhance model performance through efficient model integration demonstrates its potential as a valuable tool in medical imaging tasks.

The third approach involved employing YOLOv8 for the detection and segmentation of prostate zones. YOLOv8, one of the best iterations of the You Only Look Once (YOLO) family of models, is renowned for its exceptional speed and accuracy in object detection tasks. It utilizes advanced deep learning techniques to process images in real time, making it particularly effective for applications in medical imaging. By leveraging YOLOv8, we aimed to enhance the precision and efficiency of prostate zonal segmentation, facilitating rapid analysis and potentially improving clinical decision making.

In the subsequent sections, we delve deeper into the specifics of these methodologies and the architectural design of the developed networks.

3.2.1. Ensemble Learning

Ensemble learning stands out as a robust framework in the field of machine learning, demonstrating significant benefits across various applications. In essence, an ensemble consists of multiple individual models that operate concurrently. These models collaborate by merging their outputs through a specific decision fusion strategy to deliver a unified solution to a particular problem [24].

The core concept of ensemble learning is founded on the principle that an ensemble's generalization capability typically surpasses an individual model's capability. Over the

years, researchers have been intrigued by the reasons behind the superiority of ensemble methods compared to single learners [25]. From a technical perspective, ensemble learning primarily involves two key processes: training a set of weak component learners and strategically combining these individual learners to form a more robust overall model [24]. To create an effective final ensemble learner, researchers have developed various selection strategies to identify the most appropriate component models [26]. In this section, we provide a concise overview of several common strategies used in ensemble learning:

1. Average: This method calculates the average of outputs from different classifiers, choosing the class associated with the highest average value. It is typically used when the output from each classifier is numerical.
2. Weighted Average: Unlike the standard averaging method, this approach assigns weights to the outputs of individual classifiers based on their importance. These weights aim to minimize discrepancies between the ensemble's output and the true output, often derived from an error correlation matrix.
3. Nash Vote: in this strategy, each classifier assigns a value between zero and one for each candidate output, contributing to the decision-making process.
4. Dynamically Weighted Average: Here, weights are not static. Instead, they dynamically adjust based on the confidence levels of the outputs from the respective classifiers.
5. Weighted Average with Data-Dependent Weights: this variation in the weighted average utilizes specific partitions of the input space, calculating different weights for each partition through methods like the FSL algorithm.
6. Majority Vote: each classifier casts a vote for the class with the highest output, and the final class is the one that receives the most votes.
7. Winner Takes All (WTA): this approach selects the class with the highest output across all classifiers as the definitive class.
8. Bayesian Combination: using probabilistic approaches, this method estimates the belief value that a sample belongs to a particular class.

These strategies offer various ways to enhance the performance of ensemble learners by effectively combining the outputs of multiple classifiers [27].

Ensemble learning has proven to be a powerful approach that is widely used in segmenting anatomical structures from medical images, enhancing the accuracy and robustness of segmentation models. In this study, we employed average ensemble learning utilizing three U-Net-based models—Att-R-Net, Vanilla-Net, and V-Net—for prostate anatomical segmentation. Initially, each network was trained using 5-fold cross-validation, resulting in a total of 15 models. An average ensemble model was then constructed from these 15 models to predict prostate zones in the validation dataset. Detailed descriptions of these U-Net-based network architectures are provided in the following sections.

Attention-Res-U-Net (Att-R-Net)

Att-R-Net is an advanced neural network architecture specifically designed for medical image segmentation tasks. This model synergizes the benefits of residual blocks and attention gates to enhance segmentation accuracy and robustness. Residual blocks, which facilitate the training of deep networks by mitigating vanishing gradient issues, and attention gates, which improve the model's focus on relevant regions, are integral components of this architecture. By integrating these sophisticated elements, the Att-R-Net not only maintains high resolution in feature maps across its layers but also selectively emphasizes critical anatomical structures while suppressing irrelevant information. This results in more precise and reliable segmentation outcomes. Comprehensive details on the functioning and benefits of residual blocks and attention gates are provided in the following:

Residual Blocks: Researchers have proposed that increasing the depth of neural networks can enhance model performance. However, training deeper models often presents significant challenges, such as the issue of vanishing gradients [28]. To address these challenges, He et al. [29] introduced a deep residual learning framework that utilizes identity mapping to facilitate the training of deeper architectures. In parallel, Ronneberger et al. devised the UNet architecture, which integrates multiple-level features to improve segmentation performance. UNet has become the foundational model for biomedical image segmentation due to its ability to concatenate low-level features with higher-level ones. Building on this, Zhang et al. [28] developed ResUNet, a deeper residual U-Net that merges the strengths of both the deep residual learning strategy and UNet [30]. Incorporating residual blocks simplifies the training of deeper networks, while the model's skip connections promote efficient information flow without compromising the neural network's architecture. This results in significantly improved performance on semantic segmentation tasks while also reducing the number of parameters required [28].

Attention Gate: Attention gates are widely used in fields such as natural language processing, image analysis, and knowledge graphs. They can be categorized into two types: soft attention and hard attention [31]. Hard attention methods, such as cyclic region classifiers and pruning, are often non-differentiable and typically rely on reinforcement learning for parameter adjustment, which complicates the training of networks [32]. In contrast, soft attention is probabilistic and employs standard backpropagation instead of Monte Carlo sampling. For instance, additive soft attention is utilized in tasks like sentence rephrasing and, more recently, in image classification [31]. At the deepest level of the analysis path, the network achieves the most comprehensive feature representation. However, the use of cascaded convolutions and nonlinear activation functions can lead to the loss of spatial information in the high-level output maps [33]. Consequently, this can make it difficult to reduce false detections for smaller lesions that exhibit significant variations in size and shape. Figure 1 demonstrates the architecture of our proposed Att-R-Net.
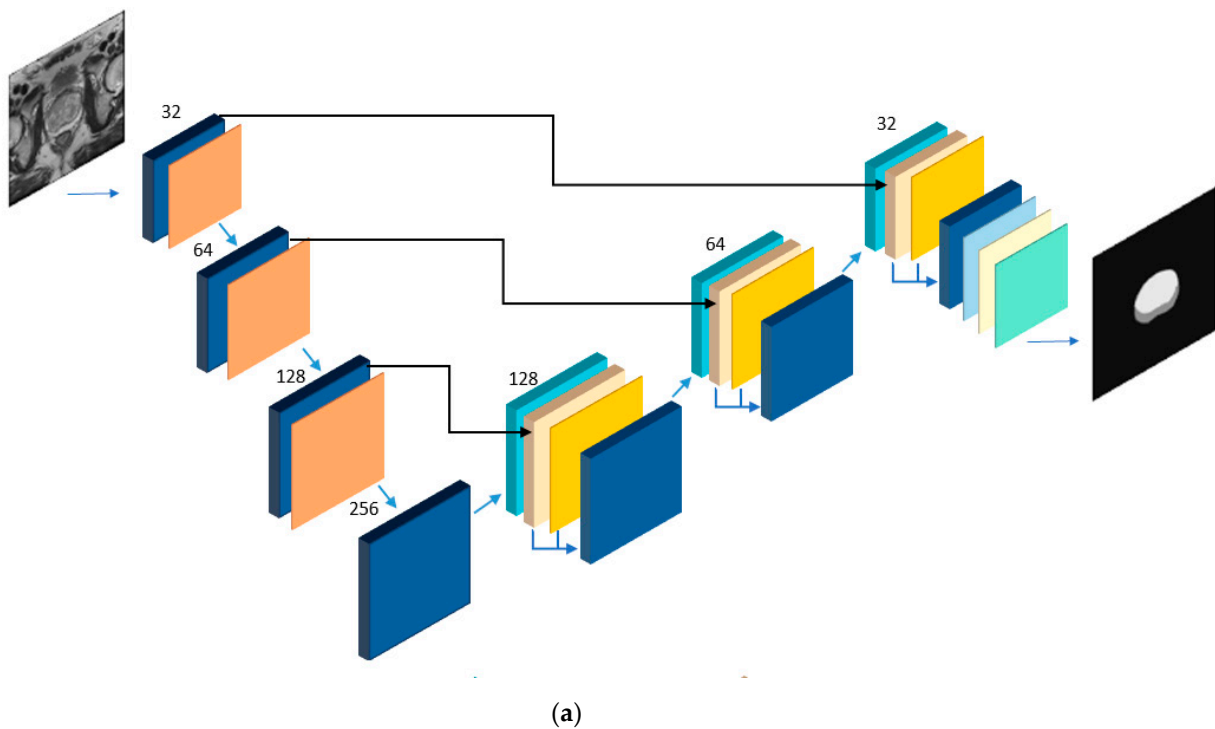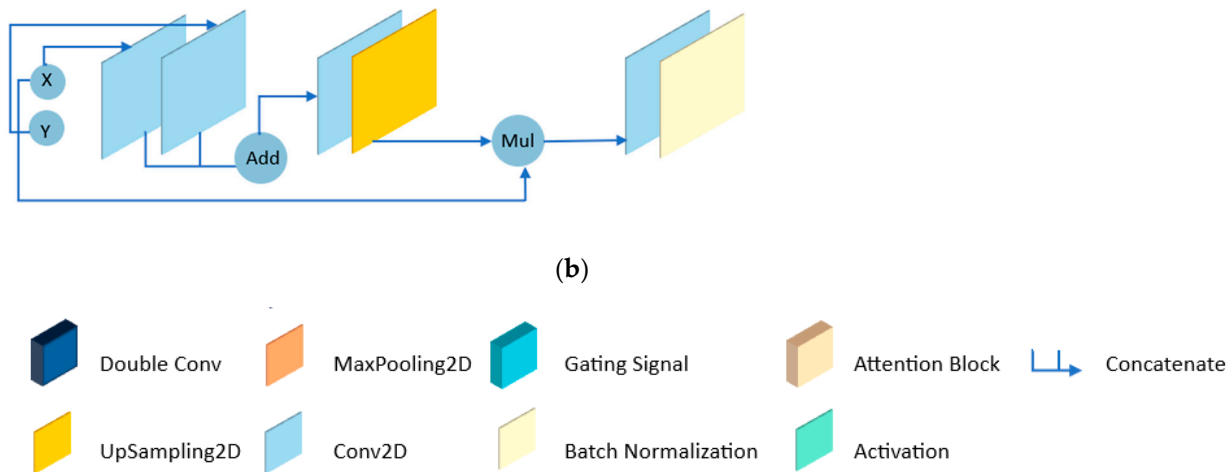

(**a**)

Figure 1. *Cont*.

(b)

Double Conv MaxPooling2D Gating Signal Attention Block Concatenate

UpSampling2D Conv2D Batch Normalization Activation

Figure 1. The architecture and details of the Att-R-Net neural network, (**a**) the overall network, and (**b**) the architecture of the attention block. The Double Conv contains two convolution layers with the Relu activation function, and the gating signal contains one convolution layer with the Relu activation function.

Vanilla-Net

Vanilla-Net stands out with its minimalist approach to neural network architecture. By forgoing deep layers, shortcuts, and complex operations like self-attention, Vanilla-Net achieves a balance of simplicity and robustness. Each layer is meticulously optimized for compactness and simplicity, with nonlinear activation functions removed post-training to maintain the original structure. This design effectively mitigates the inherent complexity of traditional networks, making Vanilla-Net a perfect fit for environments with limited resources. Its straightforward and highly simplified design paves the way for efficient deployment, without sacrificing performance. Extensive experiments reveal that Vanilla-Net performs comparably to well-known deep neural networks and vision transformers. This highlights the power of minimalism in deep learning. The pioneering approach of Vanilla-Net has the potential to reshape the landscape and challenge conventional models, setting a new standard for elegant and effective network design [34]. The structure of Vanilla-Net is illustrated in Figure 2.

The get_crop_shape function in the U-Net architecture is crucial for aligning the dimensions of feature maps from the encoder and decoder during concatenation in the upsampling process. It calculates the required cropping sizes by comparing the spatial dimensions of a target feature map from the encoder and a reference from the decoder. By addressing width and height differences, the function ensures proper alignment of the concatenated feature maps, which helps preserve spatial information and enhances the model's performance in segmentation tasks [30].

V-Net

V-Net, originally developed for 3D volumetric data segmentation, has shown considerable promise in adapting its architecture for 2D image segmentation tasks. Though primarily associated with medical imaging, the principles behind V-Net can be effectively harnessed for segmenting a wide variety of 2D images, especially in scenarios requiring precise boundary delineation. The architecture of V-Net is inspired by the widely popular U-Net model, featuring an encoder–decoder structure that incorporates skip connections. These skip connections play a crucial role in preserving high-resolution features, which are vital for achieving accurate segmentation results [30]. In its original form, V-Net utilizes 3D convolutions and is specifically designed for volumetric medical image segmentation [35].

However, adaptations to 2D convolutional layers make it suitable for traditional image datasets, allowing V-Net to effectively capture spatial hierarchies in 2D images. The model's use of a Dice coefficient-based loss function is particularly advantageous in addressing class imbalances, such as segmenting tumors or small organs [36].
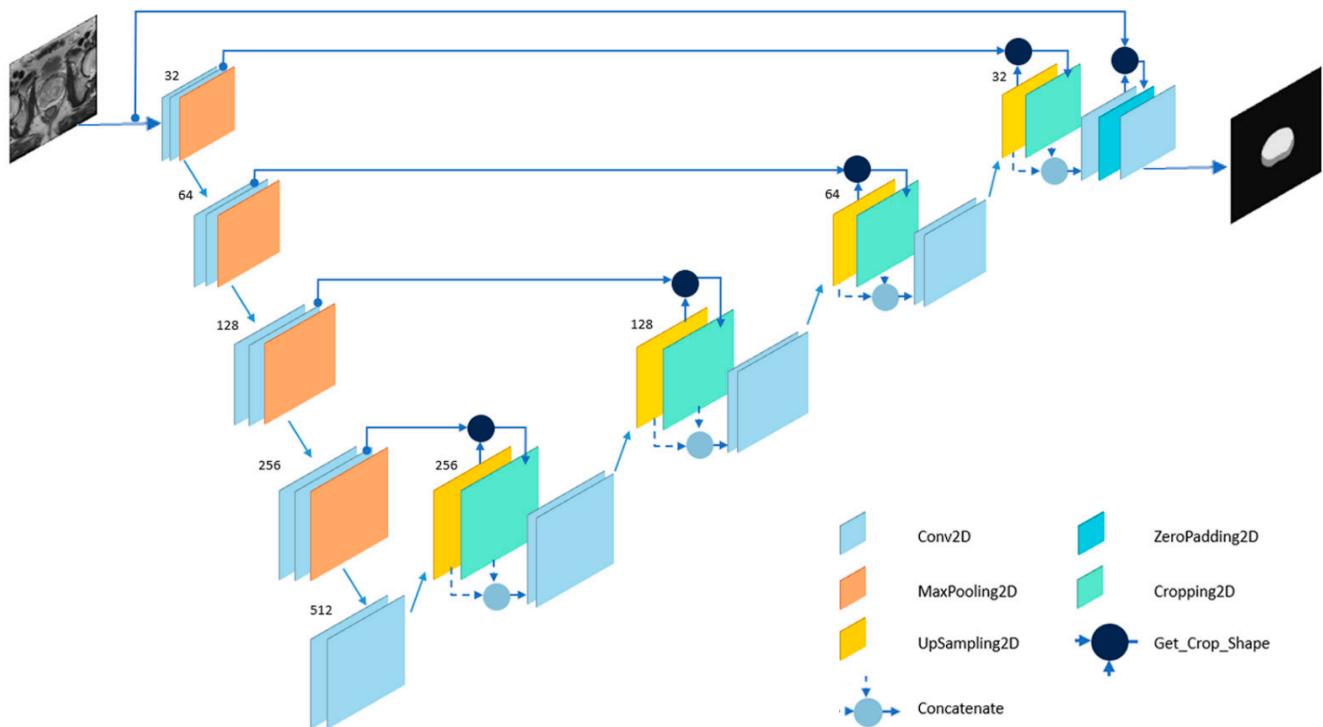

Figure 2. The architecture and details of the Vanilla-Net neural network.

Moreover, V-Net's efficient architecture enables quick training and inference, critical in real-time applications. This efficiency, combined with its competitive performance, allows V-Net to handle diverse 2D segmentation tasks, extending its utility to fields like remote sensing and natural scene understanding. For instance, it can identify land cover types in satellite imagery or segment objects in autonomous driving scenarios, showcasing its versatility across various domains. Figure 3 shows the architecture of the suggested V-Net neural network.

3.2.2. Meta-Net

MetaNet, based on the so called Theory of Independent Judges (TIJ) [37], is applied to the context of substance use and misuse. In TIJ, each artificial neural network (ANN) acts as an expert judge for the specific problem it encounters, with its credibility determined by performance during testing and/or validation. For classification problems (1 of N), each judge-ANN holds varying credibility for different aspects of the problem, with each aspect representing one of the N encoded input classes. The credibility of each judge is implicitly captured in the Confusion Matrix (CM), which details performance during the testing phase. TIJ proposes that with M judge-ANNs and their respective M Confusion Matrices, a MetaNet can be developed. This MetaNet takes as input the combined outputs of all M judges and provides the N classes as output. The weight matrix is created through an algorithm that processes the judges' CMs, and signal propagation follows a cooperative and competitive feed-forward algorithm (Figure 4) [38].
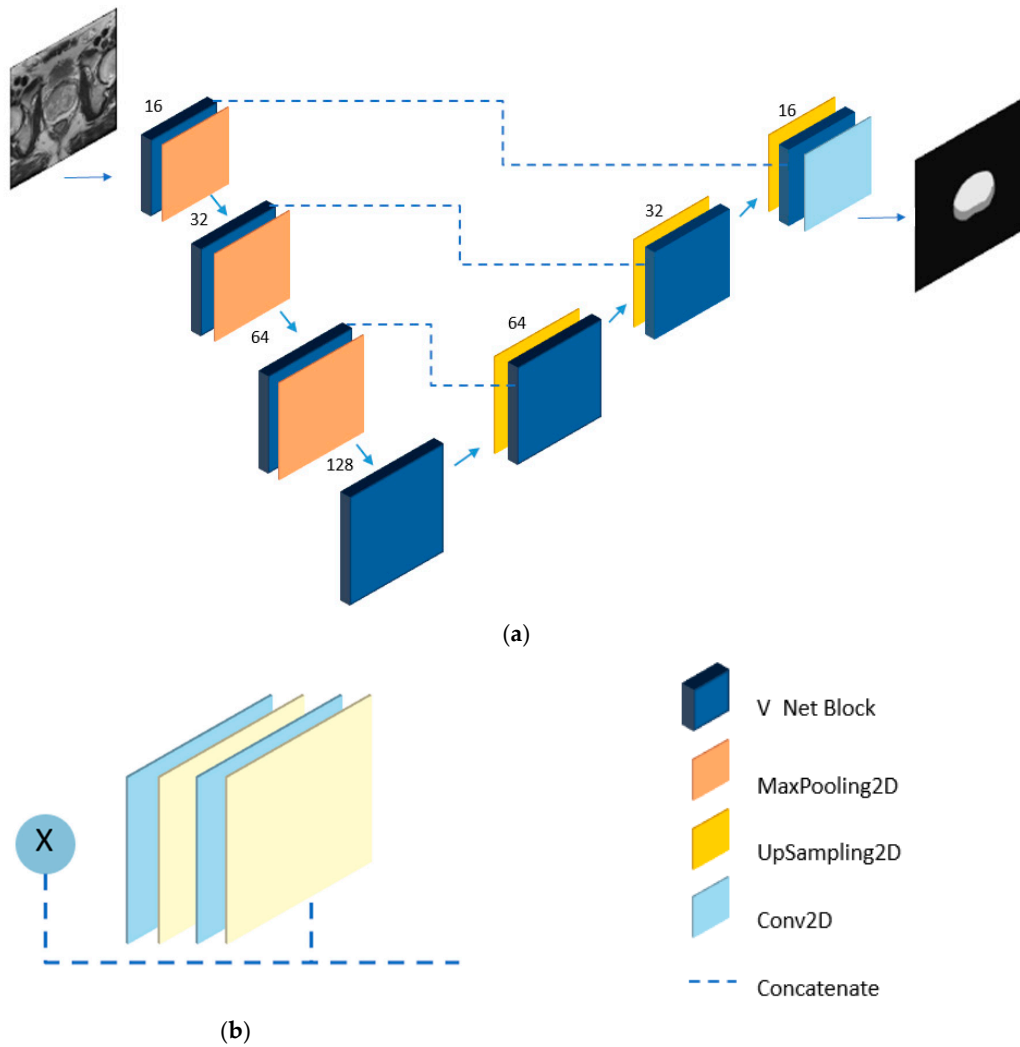
Figure 3. The architecture and details of the V-Net neural network, (**a**) the overall network, and (**b**) the architecture of the V-Net Block.
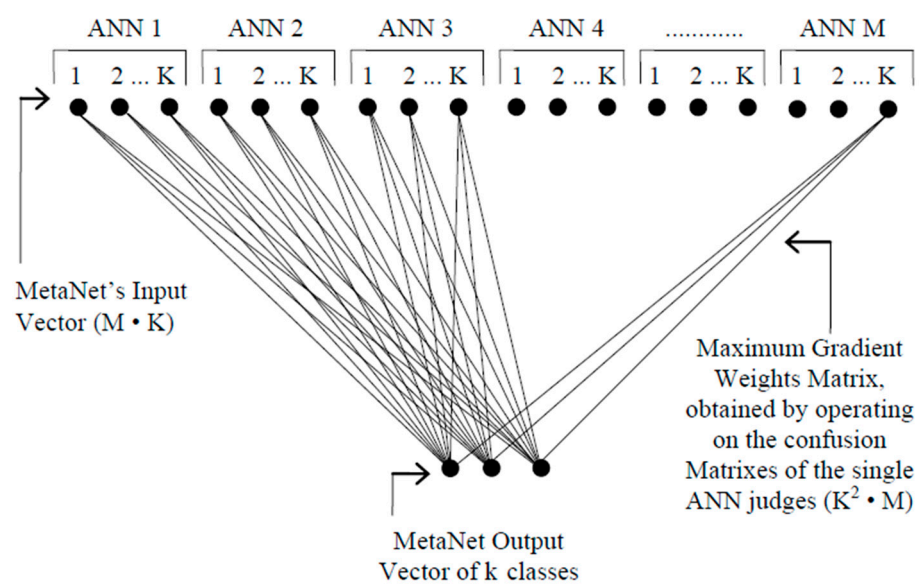

Figure 4. The structure of Meta-Net.

The purpose of MetaNet's Weights Matrix is to define the local credibility of each judge concerning the classification problem. For an ANN to join the MetaNet judges' pool, it must provide a performance history (curriculum) related to the same problem.

In this context, the CM comparing correct classification and predicted classification generated by each ANN during the testing phase serves as a valuable curriculum. The CM of each judge-ANN will be interpreted as follows below (Figure 5).

		Output $_j$						
	ANN 1	A	B	C	......	Z	*Null*	*Tot. Row*
	A	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	......	$X_{1,Z}$	$X_{1,Z+1}$	$\dot{R}_1$
	B	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	......	$X_{2,Z}$	$X_{2,Z+1}$	$\dot{R}_2$
Target $_i$	C	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	......	$X_{3,Z}$	$X_{3,Z+1}$	$\dot{R}_3$
	....	......	......	......	......	......	......	....
	Z	$X_{Z,1}$	$X_{Z,2}$	$X_{Z,3}$	......	$X_{Z,Z}$	$X_{Z,Z+1}$	$\dot{R}_z$
	Tot. Column	$\dot{C}_1$	$\dot{C}_2$	$\dot{C}_3$	......	$\dot{C}_z$	$\dot{C}_{z+1}$	

Figure 5. Confusion Matrix of each judge-ANN.

From which:

$$Class\ specific\ precision\ E_j = \frac{X_{jj}}{C_j}$$

$$Class\ specific\ recall:\ S_i = \frac{X_{ii}}{R_i}$$

$$B_i = 1 - S_i = \frac{\sum_{i \neq j}^{Z+1} X_{ij} - X_{ii}}{R_i}$$

$$F_j = 1 - E_j = \frac{\sum_{j \neq i}^{Z+1} X_{ij} - X_{jj}}{C_j}$$

Here, S_j represents the class-specific recall for class *i* (called Successes in [37]), B_j stands for $1-S_i$ (it is called Failed Blows in [37]), E_j denotes the class-specific precision for class j (called Correct Eliminations in [37]), and F_j is $1-E_j$ (it is called False Attributions in [37]).

Beginning with the CM of each judge-ANN, various methods can be employed to compute the Weights Matrix for MetaNet [37]. The weight matrix employed in this study is derived from the following formula:

$$w_{ij} = -lg_n \frac{B_j.F_i}{S_j.E_i}$$

In all previous studies where MetaNet was tested, it was applied exclusively to classification problems, where each judge-ANN contributed to deciding the final class from a set of predefined categories. However, in the context of this study, we extend the application of MetaNet to segmentation tasks, where the challenge is to classify each pixel independently rather than assigning a single label to the entire input image. This shifts the problem from global classification to pixel-wise decision making, requiring a more granular

approach to credibility and performance evaluation. Specifically, the MetaNet used in this study is MetaBayes.

3.2.3. YOLO-V8 Architecture

YOLO was introduced to the computer vision field in 2015 through a paper by Joseph Redmon and colleagues titled "You Only Look Once: Unified, Real-Time Object Detection" [39]. This work revolutionized object detection by framing it as a straightforward regression problem, transitioning from image pixels to predicting bounding boxes and class probabilities. The "unified" approach allowed for the simultaneous prediction of multiple bounding boxes and class probabilities, enhancing speed and accuracy. From its launch in 2016 until the present year (2024), the YOLO series has rapidly advanced. Although Joseph Redmon ceased his work on YOLO at version 3 [40], various researchers have further refined the foundational "unified" concept, culminating in the recent release of YOLO-v10 [41]. In this research, we implemented YOLO-V8 to segment prostate zones.

YOLO-V8 was launched in January 2023 by Ultralytics, the same team behind YOLO-v5. While a formal research paper is forthcoming and additional features are being integrated into the YOLO-v8 repository, preliminary comparisons indicate that this latest version surpasses its predecessors, establishing itself as the new state of the art in the YOLO series [42].

YOLO-V8 has achieved state-of-the-art performance by enhancing its model architecture, incorporating both anchor box and anchor-free approaches, and utilizing a wide range of data augmentation techniques. This version supports various tasks, including object segmentation, instance segmentation, and image classification, boosting its versatility for multiple applications. As the latest iteration in the YOLO object segmentation framework, YOLO-V8 focuses on improving accuracy and efficiency compared to previous versions. Key enhancements include a refined network design, a novel approach to anchor boxes, and an updated loss function, all of which contribute to markedly improved segmentation accuracy [43]. Figure 6 demonstrates the structure and architecture of the YOLO-V8 neural network.
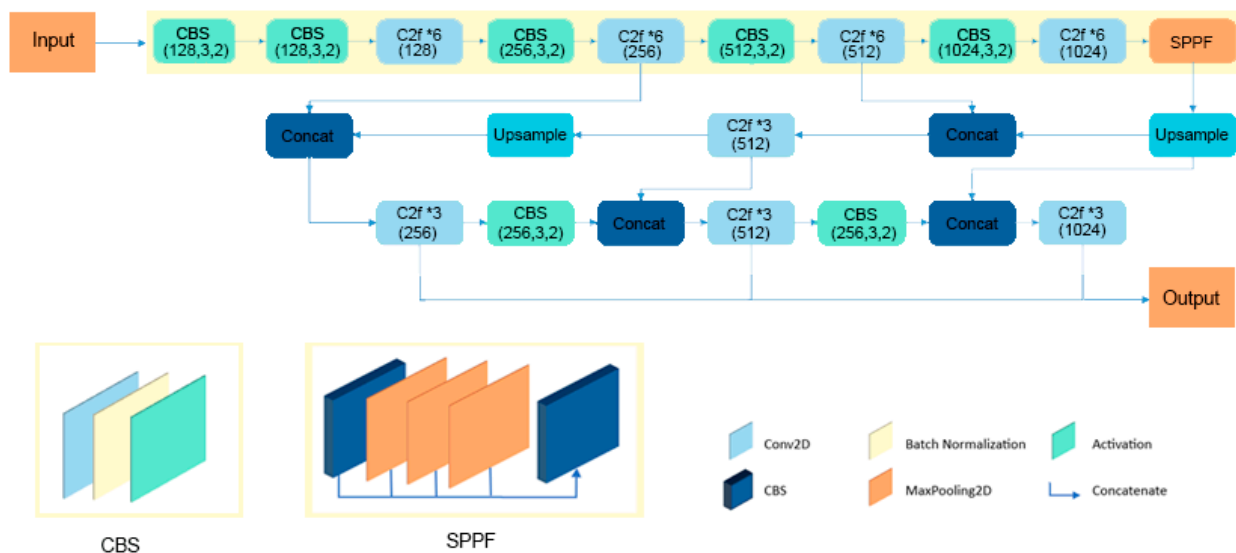

Figure 6. The diagram of the YOLOv8 network structure (figure by the authors) illustrates that the CBS component consists of convolution, batch normalization, and SiLu activation functions. Additionally, the SPPF is built from three tiers of Maxpooling integrated with two CBS units, as in [44].

3.3. Implementation Details

The current study focuses on developing three distinct methods for segmenting prostate zones (CG and PZ) using T2-weighted images. A total of 3553 T2-weighted slices from 139 multi-parametric MRI (mp-MRI) patients were retrieved from the Prostate158 public dataset. For zonal segmentation, the dataset was divided into three folds: a training set with 2306 slices from 90 patients, a validation set with 637 slices from 25 patients, and a test set with 610 slices from 24 patients.

To normalize the intensity of MRI images in all approaches, we used a custom approach based on intensity windowing. First, each image was read using SimpleITK, and we calculated the mean and standard deviation of the image intensities. The intensity normalization was then performed by adjusting the image values using the computed mean and standard deviation. Specifically, we set the intensity range to fall between one standard deviation below the mean and two standard deviations above it, mapping the pixel values within this range to a 0–255 scale. This normalization technique helped standardize image intensity across the dataset, making it suitable for further segmentation tasks.

The first approach utilizes an average ensemble of Att-R-Net, Vanilla-Net, and V-Net. All networks were trained end-to-end with the Adam optimizer, configured with an initial learning rate of 1×10^{-4}, beta_1 at 0.9, beta_2 at 0.999, and epsilon at 1×10^{-8}. The training process used a batch size of 8 and was conducted over a maximum of 300 epochs, with early stopping applied after 50 epochs if there was no improvement in validation loss. Pixel intensity was standardized to achieve a zero mean and unit variance. The T2-weighted images were resized to 144×144 pixels, and data augmentation techniques were employed during training to mitigate overfitting. The neural network's performance is significantly influenced by the choice of loss function, as it dictates how the network learns and updates its parameters based on the gradients with respect to the weights. Therefore, selecting an appropriate loss function is crucial for guiding the optimization process effectively. For training our neural networks, we employed a fixed focal loss, defined by the following formula:

$$\mathcal{L}_{focal}\left(y_{true}, y_{pred}\right) = -\alpha.(1 - p_{true})^{\gamma}.\log(p_{true} + \epsilon) - (1 - \alpha).\left(p_{false}\right)^{\gamma}.\log\left(1 - p_{false} + \epsilon\right)$$

where α is the balancing factor and is set to 0.25, γ is the focusing parameter and is set to 2, and ϵ is a small constant added for numerical stability to avoid taking the logarithm of zero. In this formula, y_{true} is the original mask, and y_{pred} is the predicted mask using our models. Also, p_{true} and p_{false} are based on the true labels y_{true}:

If $y_{true} = 1$, $p_{true} = y_{pred}$ and $p_{false} = 1 - y_{pred}$.

If $y_{true} = 0$, $p_{true} = 1 - y_{pred}$ and $p_{false} = y_{pred}$.

For YOLO, we used the GitHub repository provided by Ultralytics [45]. We converted the segmentation masks to the "YOLO format" to perform the training operation, setting a batch size of 16 using an image size of 256 and using 0.01 as the learning rate. In particular, we used the YOLO-V8 nano-segmentation version.

4. Results

All models were evaluated using intersection over union (IoU) and the DSC metrics, which were shown in Equations (1) and (2). In most segmentation studies, network performance is typically evaluated by calculating metrics across all slices of the test dataset and then averaging the results. However, in this study, we adopted a different approach by assessing performance on a patient-by-patient (case-by-case) basis. For each patient, the metrics were computed individually, reflecting a more clinically relevant evaluation method. We then calculated the median values along with the first quartile (Q1) and third quartile (Q3), providing a robust assessment of the model's performance variability across

different cases. This method offers a clearer understanding of how the network performs on a case level rather than on a slice-by-slice basis.

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \tag{1}$$

$$DSC = \frac{2|X \cap Y|}{|X \cup Y|} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

Tables 2 and 3 summarize the results of all 15 models and the average ensemble model for zonal segmentation on the validation and test sets, respectively. In each table cell, the first line shows the median and the second row demonstrates the Q1 and Q3 results.

Table 2. The results of 5-fold cross-validation of Att-R-Net, Vanilla-Net, V-Net, and average ensemble model for zonal segmentation on validation data, (the bolded numbers show the highest results).

Models	Validation Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Att-R-Net fold 1	78.4% (72–81%)	87.9% (83–89%)	54.5% (42–66%)	70.5% (59–79%)
Att-R-Net fold 2	73.2% (60–77%)	84.5% (75–87%)	54% (38–66%)	70.1% (55–80%)
Att-R-Net fold 3	73.3% (65–77%)	84.6% (78–87%)	58.3% (42–66%)	73.6% (59–79%)
Att-R-Net fold 4	75.3% (70–79%)	85.9% (82–88%)	58.2% (43–68%)	73.6% (61–81%)
Att-R-Net fold 5	70.8% (60–77%)	82.9% (75–87%)	58.1% (48–64%)	73.5% (65–78%)
Vanilla-Net fold 1	78.5% (75–82%)	87.6% (85–90%)	55.7% (48–60%)	71.5% (65–75%)
Vanilla-Net fold 2	78.3% (74–81%)	87.8% (85–89%)	57.9% (45–61%)	73.3% (62–76%)
Vanilla-Net fold 3	77.4% (74–81%)	87.3% (85–89%)	54.3% (47–62%)	70.3% (64–77%)
Vanilla-Net fold 4	79.3% (74–81%)	88.2% (85–89%)	56.4% (51–63%)	72.1% (68–77%)
Vanilla-Net fold 5	78.5% (73–81%)	87.9% (84–89%)	58.4% (46–62%)	73.8% (63–77%)
V-Net fold 1	76.7% (71–81%)	86.8% (83–90%)	56.1% (46–65%)	71.9% (63–79%)
V-Net fold 2	76.2% (73–80%)	86.3% (84–89%)	57.3% (50–62%)	72.9% (67–77%)
V-Net fold 3	76.9% (72–81%)	86.9% (83–89%)	56.7% (49–66%)	72.3% (66–79%)
V-Net fold 4	77.7% (72–81%)	87.4% (84–90%)	57.3% (49–64%)	72.8% (66–78%)
V-Net fold 5	76.8% (70–79%)	86.9% (82–88%)	58.6% (44–62%)	73.9% (61–76%)
Ensemble	**80.4%** (76–83%)	**89.1%** (86–90%)	**60.6%** (52–69%)	**75.5%** (69–82%)

Table 3. The results of 5-fold cross-validation of Att-R-Net, Vanilla-Net, V-Net, and average ensemble model for zonal segmentation on test data, (the bolded numbers show the highest results).

Models	Test Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Att-R-Net fold 1	78.9% (66–82%)	88.2% (79–90%)	49.4% (39–59%)	66.1% (56–74%)
Att-R-Net fold 2	70.6% (62–79%)	82.7% (77–88%)	47.3% (39–60%)	64.1% (56–75%)
Att-R-Net fold 3	74.5% (64–80%)	85.3% (78–89%)	49.8% (36–60%)	66.4% (56–75%)
Att-R-Net fold 4	76.5% (67–81%)	86.7% (44–62%)	52.8% (80–89%)	69.1% (61–76%)
Att-R-Net fold 5	71.7% (62–78%)	83.5% (76–88%)	49.8% (35–60%)	66.5% (52–75%)
Vanilla-Net fold 1	77.9% (67–82%)	87.6% (80–90%)	52.8% (44–58%)	69.1% (61–73%)
Vanilla-Net fold 2	78.6% (71–83%)	88% (83–90%)	53.6% (45–62%)	69.8% (62–77%)
Vanilla-Net fold 3	78.2% (69–81%)	87.8% (82–89%)	50.8% (43–59%)	67.4% (60–74%)
Vanilla-Net fold 4	77.5% (67–83%)	87.3% (80–90%)	50.1% (42–58%)	66.7% (59–73%)
Vanilla-Net fold 5	77.6% (66–82%)	87.4% (45–62%)	52.7% (80–90%)	69% (62–77%)
V-Net fold 1	75% (63–81%)	85.7% (77–89%)	48.2% (43–59%)	65.1% (60–74%)
V-Net fold 2	77% (70–81%)	87% (82–89%)	50.5% (41–59%)	67.1% (59–74%)
V-Net fold 3	75.7% (67–82%)	86.1% (80–90%)	48.2% (39–59%)	65% (56–74%)
V-Net fold 4	76.7% (68–80%)	86.8% (81–89%)	52% (44–58%)	68.4% (61–73%)
V-Net fold 5	77.1% (68–81%)	87% (81–89%)	50.2% (45–57%)	66.9% (62–73%)
Ensemble	**79.3%** (72–85%)	**88.4%** (83–92%)	**54.5%** (46–63%)	**70.5%** (63–77%)

Figure 7, which presents a comparison between the segmented masks and the coarsely annotated labels across several samples of the test set, demonstrates that the ensemble model effectively captured precise boundaries of the prostate zones.

Tables 4 and 5 present the results of the Meta-Net, which utilizes various combinations of Att-R-Net, Vanilla-Net, and V-Net for zonal segmentation on both the validation and test sets. In these tables, Att-R-Net encompasses all models from folds 1 to 5 of the network, and the same applies to Vanilla-Net and V-Net. Additionally, the notation "Att-R-Net + Vanilla-Net" indicates the combination of all attention models with Vanilla-Net models, and similar combinations are represented throughout the tables.
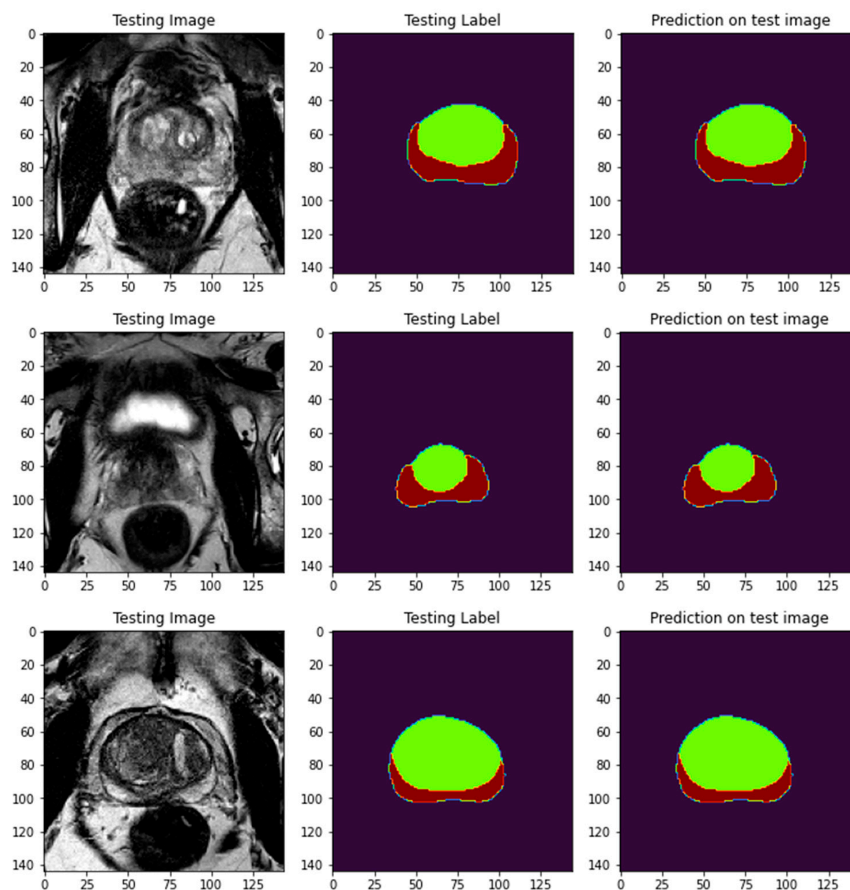
Figure 7. Segmentation results of the prostate zones using the ensemble model of three examples of the test set. Columns from left to right show images of the original image, original mask, and predicted mask of CG and PZ, (Brown: PZ and Light Green: CG).

Table 4. The results of Meta-Net using different combinations of U-net-based networks for zonal segmentation on the validation dataset, (the bolded numbers show the highest results).

Models	Validation Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Att-R-Net	75% (67–79%)	85% (80–88%)	56% (51–67%)	72% (58–80%)
Vanilla-Net	79% (76–83%)	88% (86–90%)	**58%** (49–63%)	73% (66–78%)
V-Net	78% (75–83%)	88% (86–91%)	57% (49–64%)	72% (66–78%)
Att-R-Net + Vanilla-Net	**80%** (74–83%)	**89%** (85–90%)	**58%** (47–67%)	**74%** (64–80%)
Att-R-Net + V-Net	79% (73–82%)	88% (85–90%)	56% (46–67%)	72% (63–80%)
Vanilla-Net + V-Net	**80%** (76–83%)	**89%** (87–91%)	**58%** (52–65%)	73% (68–79%)
Att-R-Net + Vanilla-Net + V-Net	**80%** (76–83%)	**89%** (86–91%)	**58%** (51–67%)	73% (67–80%)

Table 5. The results of Meta-Net using different combinations of U-net-based networks for zonal segmentation on the test dataset, (the bolded numbers show the highest results).

Models	Test Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Att-R-Net	74% (68–82%)	85% (81–90%)	49% (42–62%)	65% (59–77%)
Vanilla-Net	78% (72–84%)	**88%** (84–91%)	**54%** (45–62%)	70% (62–77%)
V-Net	78% (71–83%)	87% (83–91%)	52% (45–60%)	69% (62–75%)
Att-R-Net + Vanilla-Net	**79%** (72–85%)	**88%** (84–92%)	51% (44–62%)	68% (61–77%)
Att-R-Net + V-Net	78% (70–85%)	**88%** (82–92%)	51% (44–62%)	68% (44–62%)
Vanilla-Net + V-Net	78% (72–85%)	**88%** (84–92%)	**54%** (47–62%)	**71%** (64–77%)
Att-R-Net + Vanilla-Net + V-Net	**79%** (71–85%)	**88%** (83–92%)	53% (46–62%)	69% (63–77%)

Figure 8 illustrates some examples of zonal segmentation using Meta-Net (Vanilla-Net + V-Net) on the test set.
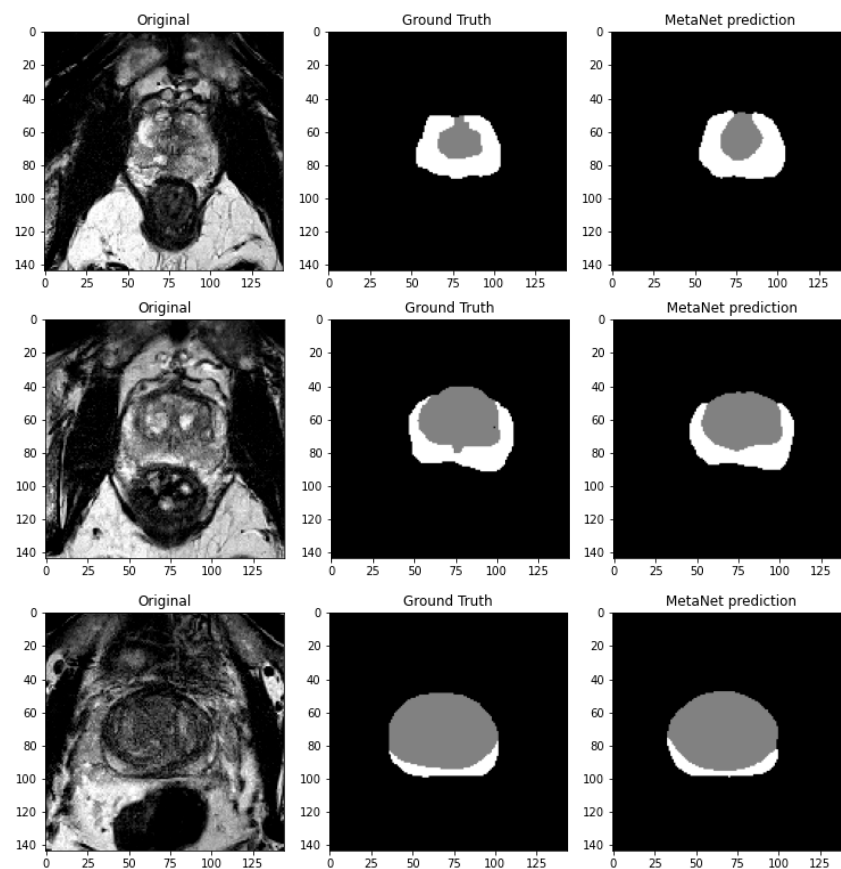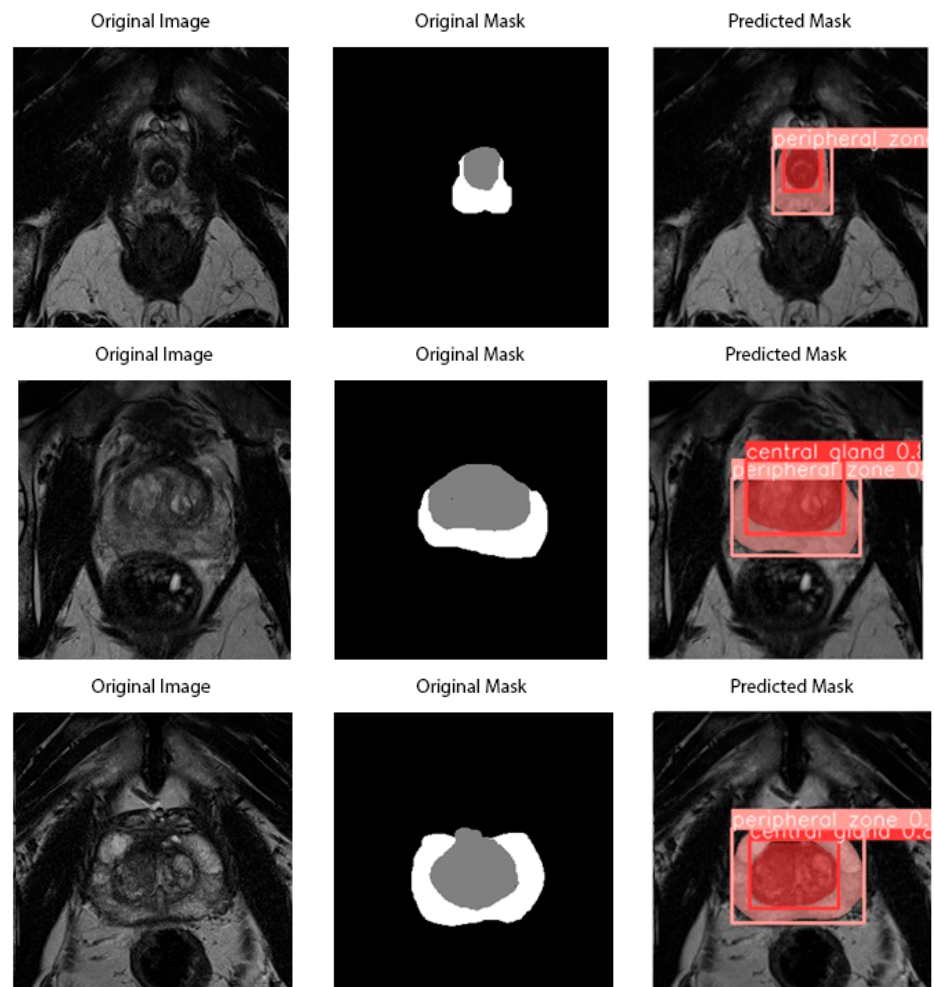

Figure 8. Segmentation results of the Meta-Net: (**left**) original image, (**middle**) ground truth, and (**right**) predicted segmentation mask.

The outcomes of YOLO-V8 are presented in Tables 6 and 7 for the validation dataset and test set. Figure 9 illustrates YOLO's zonal detection and segmentation performance on some examples from the test set.

Table 6. The results of YOLO-V8 for zonal segmentation on the validation dataset.

| Model | Validation Set | | | |
| | CG | | PZ | |
	IoU	DSC	IoU	DSC
YOLO-V8	81% (70–85%)	91% (84–93%)	60% (55–69%)	74% (67–81%)

Table 7. The results of YOLO-V8 for zonal segmentation on the test dataset.

| Model | Test Set | | | |
| | CG | | PZ | |
	IoU	DSC	IoU	DSC
YOLO-V8	80% (71–84%)	89% (83–91%)	58% (52–66%)	73% (68–79%)

Figure 9. Detection and segmentation results of the YOLO-V8: (**left**) original image, (**middle**) ground truth, and (**right**) predicted segmentation mask.

Tables 8 and 9 demonstrate the results of three proposed approaches for PZ and CG segmentation on the validation set and test set, respectively. For Meta-Net, the results of the best combination (Vanilla-Net + V-Net) are considered as the final results.

Table 8. Comparison of the results obtained from the ensemble model, Meta-Net, and YOLO-V8 for zonal segmentation on the validation set.

Models	Validation Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Ensemble Model	80.4% (76–83%)	89.1% (86–90%)	60.6% (52–69%)	75.5% (69–82%)
Meta-Net (Vanilla-Net + V-Net)	80% (76–83%)	89% (87–91%)	58% (52–65%)	73% (68–79%)
YOLO-V8	81% (70–85%)	91% (84–93%)	60% (55–69%)	74% (67–81%)

Table 9. Comparison of the results obtained from the ensemble model, Meta-Net, and YOLO-V8 for zonal segmentation on the test set.

Models	Test Set			
	CG		PZ	
	IoU	DSC	IoU	DSC
Ensemble Model	79.3% (72–85%)	88.4% (83–92%)	54.5% (46–63%)	70.5% (63–77%)
Meta-Net (Vanilla-Net + V-Net)	78% (72–85%)	88% (84–92%)	54% (47–62%)	71% (64–77%)
YOLO-V8	80% (71–84%)	89% (83–91%)	58% (52–66%)	73% (68–79%)

The performance of the three models—ensemble model, Meta-Net (Vanilla-Net + V-Net), and YOLO-V8—was evaluated on both the CG and PZ using IoU and DSC metrics. In the CG region, YOLO-V8 achieved the best results with an IoU of 80% and DSC of 89%, followed by the ensemble model with an IoU of 79.3% and DSC of 88.4% and Meta-Net with an IoU of 78% and DSC of 88%. For the PZ region, YOLO-V8 again outperformed the other models with an IoU of 58% and DSC of 73%, while the ensemble model and Meta-Net had comparable results, with IoUs of 54.5% and 54% and DSCs of 70.5% and 71%, respectively. Figure 10 presents a comparative analysis of the IoU and DSC results obtained from the test set.
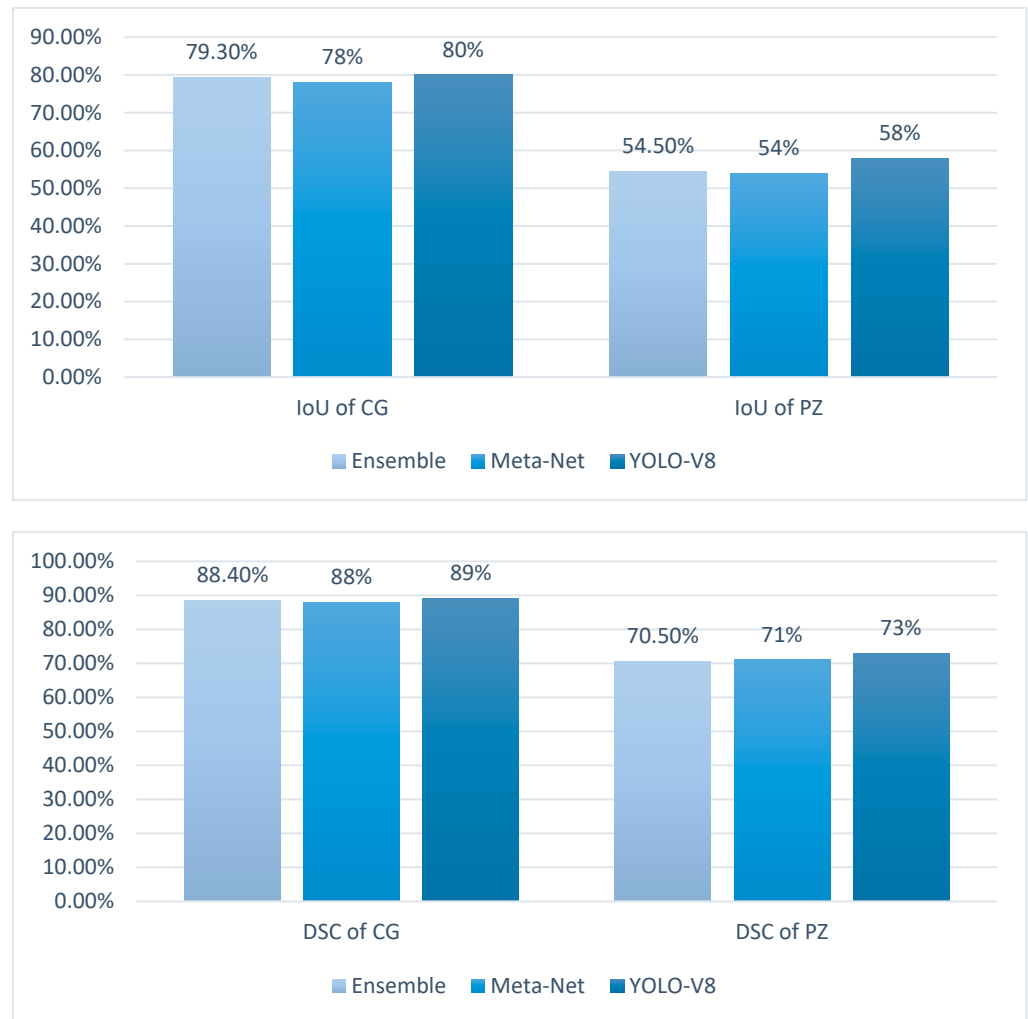
Figure 10. Comparison of the IoU and DSC results obtained from the test set.

5. Discussion

Segmentation of prostate zones, particularly the PZ, plays a vital role in the diagnosis and management of prostate cancer. This study explores three state-of-the-art neural network-based methods—ensemble learning, MetaNet, and YOLO-V8—specifically tailored to overcome the complexities of zonal segmentation in prostate MRI scans from the Prostate158 dataset. Each approach offers distinct advantages, yet they share a common aim: to elevate the precision and reliability of segmentation, thereby enhancing clinical decision making and improving patient outcomes.

The ensemble learning method used in this study, which integrates the strengths of multiple U-Net-based models, demonstrated effectiveness by averaging the outputs of different models to enhance the segmentation of both the CG and the PZ. This emphasizes the value of utilizing diverse models to account for anatomical variations in prostate MRI data.

The use of MetaNet, specifically tailored for segmentation, is a novel approach. Its strong performance highlights the potential of advanced architectures to manage segmentation tasks by capturing complex feature relationships in medical images. Although typically employed for different purposes, MetaNet's success here suggests opportunities for further exploration of state-of-the-art neural networks in medical imaging.

YOLO-V8 also showed promising results, highlighting its capability for fast and accurate segmentation. YOLO's architectural improvements, optimized for segmentation rather than traditional object detection, provide a flexible solution with potential applica-

tions across various medical imaging fields. Despite its common use in object detection, the adaptation of YOLO for zonal segmentation demonstrates its versatility and broad applicability.

The networks were trained using a dataset of 90 MRIs, with performance evaluated on a validation set of 25 MRIs and a test set of 24 MRIs. For CG segmentation on the validation set, the ensemble learning method achieved an IoU of 80.4% and a DSC of 89.1%. For PZ segmentation, it attained an IoU of 60.6% and a DSC of 75.5%. On the test set, the ensemble method achieved an IoU of 79.3% and a DSC of 88.4% for CG segmentation and an IoU of 54.5% and DSC of 70.5% for PZ segmentation.

In Meta-Net, various network combinations were evaluated on both the validation and test sets. Among these, the combination of Vanilla-Net + V-Net demonstrated the best performance. This configuration achieved an IoU of 80% and 58% for CG segmentation, along with a DSC of 89% and 73% for PZ segmentation on the validation set. On the test set, it maintained strong results with an IoU of 78% and 54% for CG segmentation and a DSC of 88% and 71% for PZ segmentation.

YOLO-V8 achieved an IoU of 81% for CG segmentation and 60% for PZ segmentation on the validation set, with corresponding DSC values of 91% and 74%. However, on the test set, its performance significantly improved, achieving an IoU of 80% for CG segmentation and 58% for PZ segmentation, along with DSC values of 89% and 73%, respectively.

In comparing our work with related studies, it is essential to highlight that our dataset is smaller than those utilized in the majority of the referenced studies, which inevitably impacts the performance metrics. Nonetheless, our results demonstrate competitive efficacy in prostate zone segmentation, marking a significant achievement given the constraints of our data. For example, ENet achieved impressive DSC scores of 91% for the whole gland, 87% for the TZ, and 71% for the PZ in the study [12]. In contrast, our ensemble learning method attained a DSC of 89.1% for the CG and 75.5% for the PZ on the validation set, showcasing particularly strong performance in CG segmentation.

When examining the results of BASC-Net [13], which reported DSC scores of 88.6% for CG and 79.9% for PZ, it becomes evident that our method closely aligns with CG performance while slightly lagging in PZ segmentation. This is particularly noteworthy considering the size and diversity of the datasets used in these studies, suggesting that our approach is robust and capable of yielding reliable results even with fewer training samples. Similarly, Meta-Net demonstrated impressive performance with an IoU of 80% for CG segmentation and a DSC of 89%. This further underscores the competitiveness of our method, particularly in CG segmentation.

The ensemble model in [14] achieved a DSC of 91.5% for the prostate gland, 86.5% for the TZ, and 73.6% for the PZ. Our ensemble approach showed comparable performance in CG segmentation (88.4% DSC) but slightly lower for PZ (70.5% DSC on the test set). The use of pre-training and a revised loss function in [14] may have contributed to its higher performance.

A 3D U-Net-based approach in [15] reported DSC values of 90.9% for CG and 84.4% for the PZ across different datasets. Our methods performed competitively in CG segmentation, with YOLO-V8 achieving 89% DSC. However, the PZ DSC values in our study remained slightly lower (maximum of 74%), potentially due to dataset variations and differences in MRI protocols.

The CCT-U-Net model [16] leveraged Transformers and convolutional networks, achieving a DSC of 87.49% for the TZ and 80.39% for the PZ. While our CG segmentation results are in line with these findings, our PZ segmentation scores are slightly lower, highlighting the advantage of hybrid architectures in capturing complex prostate structures.

A heterogeneous dataset approach in [17] trained a U-Net-based model across multiple vendors and scanners, achieving DSC values of 88% for CG, 85% for TZ, and 72% for PZ. Our models performed similarly for CG but showed a slight drop in PZ segmentation, which may indicate a need for domain adaptation techniques when dealing with different scanner variations.

The automated ML approach in [18] obtained a DSC of 81% for TZ and 62% for PZ, which aligns closely with our PZ segmentation performance. This suggests that our models are on par with other ML-based approaches, even though our dataset size is smaller.

The PPZ-SegNet model in [19], designed as a 2D–3D ensemble, reported average DSC scores of 86% for CG and 79% for PZ. Our models demonstrated higher CG segmentation accuracy but slightly lower PZ performance, emphasizing the potential benefits of incorporating a 3D spatial context.

Dense U-Net [20] outperformed standard U-Net, reaching a DSC of 92.1% for CG and 78.1% for PZ. Our CG results are comparable, but PZ performance remains slightly lower, which may suggest the need for improved feature fusion mechanisms.

When comparing our results to those reported in [21], it is evident that their U-Net-based approach achieved higher DSC values (93.75% for CG and 89.89% for PZ on T2W images). In contrast, our best-performing model, YOLO-V8, achieved DSC values of 91% for CG and 74% for PZ on the validation set and 89% for CG and 73% for PZ on the test set. It is important to note that the authors in [21] utilized a larger dataset (225 MRIs compared to our 90 for training), which likely contributed to their higher performance. Despite the smaller dataset, our approach demonstrates competitive performance, particularly for CG segmentation, and introduces novel methodologies such as ensemble learning and Meta-Net, which have not been previously applied to zonal segmentation.

Additionally, our findings contribute to the growing body of literature on effective segmentation techniques in prostate imaging. Despite the limitations posed by a smaller dataset, our results illustrate that significant advancements can still be achieved. The ability to achieve such outcomes reinforces the idea that high-quality segmentation does not solely depend on large datasets but can also be driven by innovative model architectures and training strategies.

Overall, while our results are on par with some of the best performances reported in the literature, they highlight the feasibility of achieving strong outcomes in prostate segmentation with limited resources. This is a critical observation for future research, suggesting that further investigations could explore optimizing segmentation techniques without relying heavily on extensive datasets. Our study not only demonstrates the potential of our methods but also paves the way for ongoing research into effective prostate imaging solutions, especially in contexts where data availability may be a concern. Figures 11 and 12 present a comparison between the results of our proposed models and those from previous studies for CG and PZ segmentation, respectively. These figures display the performance of our models on the test set, with the comparison primarily based on DSC, as most prior studies did not report IoU.

Furthermore, in our study, Meta-Net model selection was performed randomly, meaning we did not incorporate a systematic criterion to choose the best-performing architecture. While this approach allowed for flexibility, it may not ensure optimal generalization. Bayesian model-selection methods, such as the Bayesian information criterion (BIC) and Bayes factors, provide a principled framework to compare models by balancing accuracy and complexity [46]. Integrating such techniques into Meta-Net could help in selecting the most suitable model while avoiding overfitting. Moreover, Bayesian deep learning approaches, like Bayesian neural networks, could quantify uncertainty in segmentation

outcomes. Future work can explore Bayesian methods to improve robustness and reliability in prostate zonal segmentation.
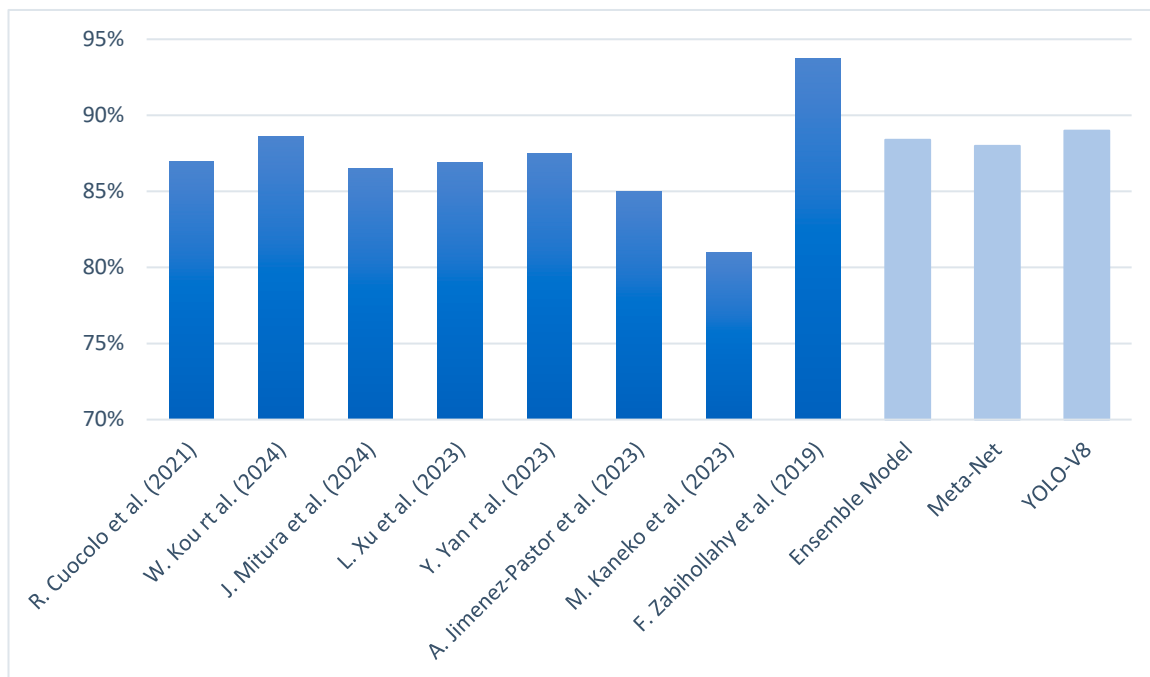

Figure 11. Comparison of the DSC results for CG segmentation, obtained from related works and model of our study on the test set [12–18,21].
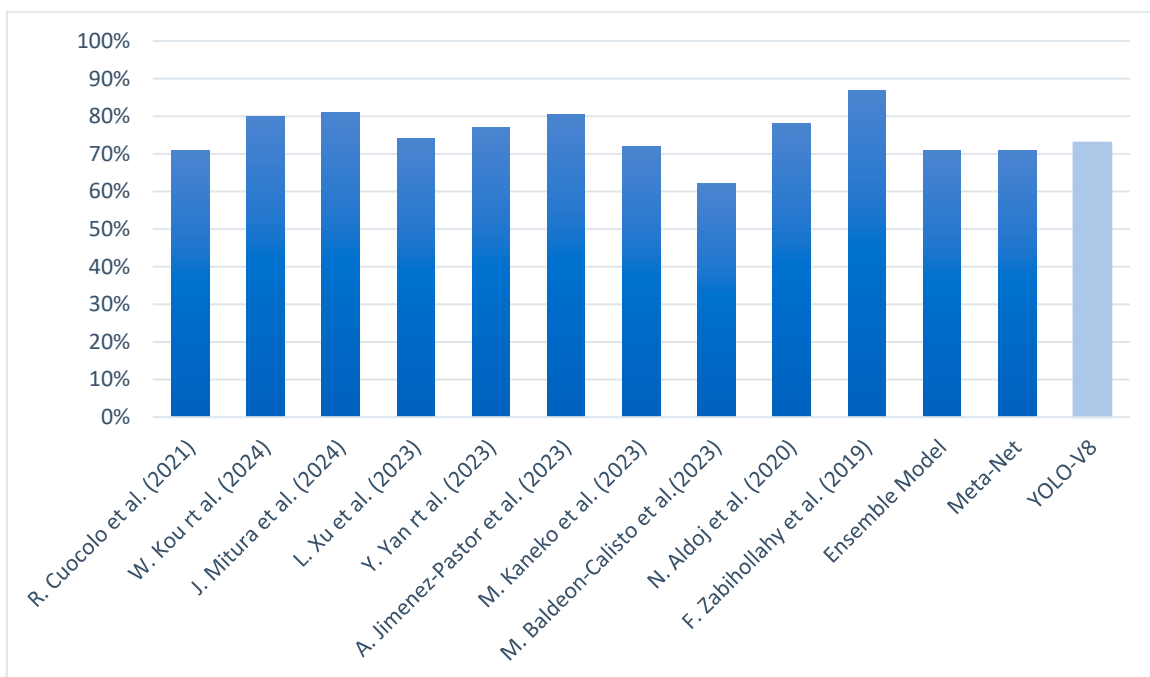

Figure 12. Comparison of the DSC results for PZ segmentation, obtained from related works and model of our study on the test set [12–21].

6. Conclusions

In conclusion, this study demonstrates the effectiveness of ensemble learning, Meta-Net, and YOLO-V8 for prostate zone segmentation. Despite the relatively small dataset compared to other studies, the models performed competitively, with YOLO-V8 showing

the best results overall, particularly for PZ segmentation. The ensemble model and Meta-Net also delivered strong results, marking the first application of Meta-Net in segmentation tasks. Notably, our findings suggest that significant improvements in segmentation performance can be achieved even with limited data, reinforcing the potential of innovative model architectures.

Author Contributions: S.F.: conceived the study, planned the experiments, wrote the manuscript, and implemented neural networks. L.D.P.: conceived the study, planned the experiments, and implemented neural networks. F.D.: contributed to data preparation. D.F.: contributed to data preparation. A.M.: revised the initial draft and contributed to the final version of the submitted manuscript. S.P.: revised the initial draft and contributed to the final version of the submitted manuscript. G.G.: revised the initial draft and contributed to the final version of the submitted manuscript. M.A.: conceived the study, planned the experiments, and wrote the initial draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Approval for this study was obtained from the Institutional Review Board following a comprehensive review process, ensuring that all ethical considerations and privacy concerns were adequately addressed to protect participants' rights and well-being.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are freely available at the following link: https://zenodo.org/records/6481141 accessed on 6 January 2025.

Conflicts of Interest: Authors Alessandro Maiocchi and Marco Alì were employed by the company R&D, Bracco Imaging S.p.a. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chahal, E.S.; Patel, A.; Gupta, A.; Purwar, A.; G, D. Unet Based Xception Model for Prostate Cancer Segmentation from MRI Images. *Multimed. Tools Appl.* **2022**, *81*, 37333–37349. [CrossRef]
2. Yang, X.; Liu, C.; Wang, Z.; Yang, J.; Le Min, H.; Wang, L.; Cheng, K.-T.T. Co-Trained Convolutional Neural Networks for Automated Detection of Prostate Cancer in Multi-Parametric MRI. *Med. Image Anal.* **2017**, *42*, 212–227. [CrossRef] [PubMed]
3. William, K.; Oh, M.D.; Mark Hurwitz, M.D.; Anthony, V.; D'Amico, M.D.; Jerome, P.; Richie, M.D.; Philip, W.; Kantoff, M. Biology of Prostate Cancer. In *Holland-Frei Cancer Medicine*, 6th ed.; National Library of Medicine: Rockville, MD, USA, 2003.
4. Adler, D.; Lindstrot, A.; Ellinger, J.; Rogenhofer, S.; Buettner, R.; Perner, S.; Wernert, N. The Peripheral Zone of the Prostate Is More Prone to Tumor Development than the Transitional Zone: Is the ETS Family the Key? *Mol. Med. Rep.* **2012**, *5*, 313–316. [CrossRef] [PubMed]
5. Holder, K.G.; Galvan, B.; Knight, A.S.; Ha, F.; Collins, R.; Weaver, P.E.; Brandi, L.; de Riese, W.T. Possible Clinical Implications of Prostate Capsule Thickness and Glandular Epithelial Cell Density in Benign Prostate Hyperplasia. *Investig. Clin. Urol.* **2021**, *62*, 423–429. [CrossRef]

6. Sato, S.; Kimura, T.; Onuma, H.; Egawa, S.; Takahashi, H. Transition Zone Prostate Cancer Is Associated with Better Clinical Outcomes than Peripheral Zone Cancer. *BJUI Compass* **2021**, *2*, 169–177. [CrossRef] [PubMed]

7. Wu, C.; Montagne, S.; Hamzaoui, D.; Ayache, N.; Delingette, H.; Renard-Penna, R. Automatic Segmentation of Prostate Zonal Anatomy on MRI: A Systematic Review of the Literature. *Insights Imaging* **2022**, *13*, 202. [CrossRef] [PubMed]

8. Korsager, A.S.; Fortunati, V.; van der Lijn, F.; Carl, J.; Niessen, W.; Østergaard, L.R.; van Walsum, T. The Use of Atlas Registration and Graph Cuts for Prostate Segmentation in Magnetic Resonance Images. *Med. Phys.* **2015**, *42*, 1614–1624. [CrossRef]

9. Nai, Y.; Teo, B.; Tan, N.; Chua, K.; Wong, C.; O'Doherty, S.; Stephenson, M.; Schaefferkoetter, J.; Thian, Y.; Chiong, E.; et al. Evaluation of Multimodal Algorithms for the Segmentation of Multiparametric MRI Prostate Images. *Comput. Math. Methods Med.* **2020**, *2020*, 8861035. [CrossRef]

10. Zaridis, D.I.; Mylona, E.; Tachos, N.; Kalantzopoulos, C.N.; Marias, K.; Tsiknakis, M.; Matsopoulos, G.K.; Koutsouris, D.D.; Fotiadis, D.I. ResQu-Net: Effective Prostate's Peripheral Zone Segmentation Leveraging the Representational Power of Attention-Based Mechanisms. *Biomed. Signal Process. Control* **2024**, *93*, 106187. [CrossRef]

11. Zavala-Romero, O.; Breto, A.L.; Xu, I.R.; Chang, Y.-C.C.; Gautney, N.; Pra, A.D.; Abramowitz, M.; Pollack, A.; Stoyanova, R. Segmentation of Prostate and Prostate Zones Using Deep Learning. *Strahlenther. Onkol.* **2020**, *196*, 932–942. [CrossRef]

12. Cuocolo, R.; Comelli, A.; Stefano, A.; Benfante, V.; Dahiya, N.; Stanzione, A.; Castaldo, A.; De Lucia, D.R.; Yezzi, A.; Imbriaco, M. Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset. *J. Magn. Reson. Imaging* **2021**, *54*, 452–459. [CrossRef] [PubMed]

13. Kou, W.; Marshall, H.; Chiu, B. Boundary-Aware Semantic Clustering Network for Segmentation of Prostate Zones from T2-Weighted MRI. *Phys. Med. Biol.* **2024**, *69*, 175009. [CrossRef]

14. Mitura, J.; Jóźwiak, R.; Mycka, J.; Mykhalevych, I.; Gonet, M.; Sobecki, P.; Lorenc, T.; Tupikowski, K. Ensemble Deep Learning Models for Segmentation of Prostate Zonal Anatomy and Pathologically Suspicious Areas BT. In *Medical Image Understanding and Analysis*; Yap, M.H., Kendrick, C., Behera, A., Cootes, T., Zwiggelaar, R., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 217–231.

15. Xu, L.; Zhang, G.; Zhang, D.; Zhang, J.; Zhang, X.; Bai, X.; Chen, L.; Peng, Q.; Jin, R.; Mao, L.; et al. Development and Clinical Utility Analysis of a Prostate Zonal Segmentation Model on T2-Weighted Imaging: A Multicenter Study. *Insights Imaging* **2023**, *14*, 44. [CrossRef] [PubMed]

16. Yan, Y.; Liu, R.; Chen, H.; Zhang, L.; Zhang, Q. CCT-Unet: A U-Shaped Network Based on Convolution Coupled Transformer for Segmentation of Peripheral and Transition Zones in Prostate MRI. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 4341–4351. [CrossRef] [PubMed]

17. Jimenez-Pastor, A.; Lopez-Gonzalez, R.; Fos-Guarinos, B.; Garcia-Castro, F.; Wittenberg, M.; Torregrosa-Andrés, A.; Marti-Bonmati, L.; Garcia-Fontes, M.; Duarte, P.; Gambini, J.P.; et al. Automated Prostate Multi-Regional Segmentation in Magnetic Resonance Using Fully Convolutional Neural Networks. *Eur. Radiol.* **2023**, *33*, 5087–5096. [CrossRef] [PubMed]

18. Kaneko, M.; Cacciamani, G.E.; Yang, Y.; Magoulianitis, V.; Xue, J.; Yang, J.; Liu, J.; Lenon, M.S.L.; Mohamed, P.; Hwang, D.H.; et al. MP09-05 Automated Prostate Gland and Prostate Zones Segmentation Using a Novel Mri-Based Machine Learning Framework and Creation of Software Interface for Users Annotation. *J. Urol.* **2023**, *209*, e105. [CrossRef]

19. Baldeon-Calisto, M.; Wei, Z.; Abudalou, S.; Yilmaz, Y.; Gage, K.; Pow-Sang, J.; Balagurunathan, Y. A Multi-Object Deep Neural Network Architecture to Detect Prostate Anatomy in T2-Weighted MRI: Performance Evaluation. *Front. Nucl. Med.* **2023**, *2*, 1083245. [CrossRef]

20. Aldoj, N.; Biavati, F.; Michallek, F.; Stober, S.; Dewey, M. Automatic Prostate and Prostate Zones Segmentation of Magnetic Resonance Images Using DenseNet-like U-Net. *Sci. Rep.* **2020**, *10*, 14315. [CrossRef]

21. Zabihollahy, F.; Schieda, N.; Krishna Jeyaraj, S.; Ukwatta, E. Automated Segmentation of Prostate Zonal Anatomy on T2-Weighted (T2W) and Apparent Diffusion Coefficient (ADC) Map MR Images Using U-Nets. *Med. Phys.* **2019**, *46*, 3078–3090. [CrossRef] [PubMed]

22. 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. *Comput. Biol. Med.* **2022**, *148*, 105817. [CrossRef]

23. Chen, P. MetaNet: Network Analysis for Omics Data. Available online: https://github.com/Asa12138/MetaNet (accessed on 6 January 2025).

24. Huang, F.; Xie, G.; Xiao, R. Research on Ensemble Learning. In Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, China, 7–8 November 2009; Volume 3, pp. 249–252.

25. Thomas, G.D. Machine Learning Research: Four Current Directions. *Artif. Intell. Mag.* **1997**, *18*, 97–136.

26. Zhou, Z.-H.; Wu, J.; Tang, W. Ensembling Neural Networks: Many Could Be Better than All. *Artif. Intell.* **2002**, *137*, 239–263. [CrossRef]

27. Verikas, A.; Lipnickas, A.; Malmqvist, K.; Bacauskiene, M.; Gelzinis, A. Soft Combination of Neural Classifiers: A Comparative Study. *Pattern Recognit. Lett.* **1999**, *20*, 429–444. [CrossRef]

28. Zhang, C.; Song, Y.; Liu, S.; Lill, S.; Wang, C.; Tang, Z.; You, Y.; Gao, Y.; Klistorner, A.; Barnett, M.; et al. *MS-GAN: GAN-Based Semantic Segmentation of Multiple Sclerosis Lesions in Brain Magnetic Resonance Imaging*; IEEE: Piscataway, NJ, USA, 2018.

29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

30. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation BT. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

31. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Y Hammerla, N.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]

32. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent Models of Visual Attention. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; Volume 27.

33. Abraham, N.; Khan, N. A Novel Focal Tversky Loss Function with Improved Attention U-Net for Lesion Segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 683–687.

34. Chen, H.; Wang, Y.; Guo, J.; Tao, D. VanillaNet: The Power of Minimalism in Deep Learning. In Proceedings of the Advances in Neural Information Processing Systems; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc.: Newry, UK, 2023; Volume 36, pp. 7050–7064.

35. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.

36. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation BT. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016.*; Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 424–432.

37. Buscema, M. MetaNet: The Theory of Independent Judges. *Subst. Use Misuse* **1998**, *33*, 439–461. [CrossRef]

38. Buscema, M. Self-Reflexive Networks. *Subst. Use Misuse* **1998**, *33*, 409–438. [CrossRef] [PubMed]

39. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

40. Zhao, L.; Li, S. Object Detection Algorithm Based on Improved YOLOv3. *Electronics* **2020**, *9*, 537. [CrossRef]

41. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Mach. Tool.* **2023**, *11*, 677. [CrossRef]

42. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. Available online: https://github.com/ultralytics/ultralytics (accessed on 12 January 2023).

43. Sahafi, A.; Koulaouzidis, A.; Lalinia, M. Polypoid Lesion Segmentation Using YOLO-V8 Network in Wireless Video Capsule Endoscopy Images. *Diagnostics* **2024**, *14*, 474. [CrossRef]

44. Guo, J.; Lou, H.; Chen, H.; Liu, H.; Gu, J.; Bi, L.; Duan, X. A New Detection Algorithm for Alien Intrusion on Highway. *Sci. Rep.* **2023**, *13*, 10667. [CrossRef]

45. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; NanoCode012; Kwon, Y.; Michael, K.; Xie, T.; Fang, J.; Yifu, Z.; et al. Ultralytics/Yolov5: V7.0—YOLOv5 SOTA Realtime Instance Segmentation. 2022. Available online: https://ui.adsabs.harvard.edu/abs/2022zndo...3908559J/abstract (accessed on 6 January 2025).

46. Thapa, S.; Lomholt, M.A.; Krog, J.; Cherstvy, A.G.; Metzler, R. Bayesian Analysis of Single-Particle Tracking Data Using the Nested-Sampling Algorithm: Maximum-Likelihood Model Selection Applied to Stochastic-Diffusivity Data. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29018–29037. [CrossRef] [PubMed]