

Cross-Slice Attention and Evidential Critical Loss for Uncertainty-Aware Prostate Cancer Detection

Alex Ling Yu Hung^{1,2}[0000–0002–2664–7430], Haoxin Zheng^{1,2}[0000–0002–8081–1940],
Kai Zhao¹[0000–0002–2496–0829], Kaifeng Pang^{1,3}[0009–0001–7613–6559], Demetri
Terzopoulos^{2,4}, Kyunghyun Sung¹[0000–0003–4175–5322]

¹ Department of Radiological Science, UCLA

² Computer Science Department, UCLA

³ Electrical and Computer Engineering Department, UCLA

⁴ VoxelCloud, Inc.

alexhung96@ucla.edu

Abstract. Current deep learning-based models typically analyze medical images in either 2D or 3D albeit disregarding volumetric information or suffering sub-optimal performance due to the anisotropic resolution of MR data. Furthermore, providing an accurate uncertainty estimation is beneficial to clinicians, as it indicates how confident a model is about its prediction. We propose a novel 2.5D cross-slice attention model that utilizes both global and local information, along with an evidential critical loss, to perform evidential deep learning for the detection in MR images of prostate cancer, one of the most common cancers and a leading cause of cancer-related death in men. We perform extensive experiments with our model on two different datasets and achieve state-of-the-art performance in prostate cancer detection along with improved epistemic uncertainty estimation. The implementation of the model is available at https://github.com/aL3x-0-o-Hung/GLCSA_ECLoss.

Keywords: MRI · Prostate Cancer · Uncertainty · Attention

1 Introduction

Prostate cancer (PCa) is one of the most common cancers and a leading cause of cancer-related death in men [22]. Multi-parametric magnetic resonance imaging (mpMRI) is commonly used for PCa diagnosis before conducting a prostate biopsy [1]. General 2D approaches such as UNet [23] and Swin-UNet [3] along with PCa detection specific approaches such as FocalNet [4] and ProstAttentionNet [8] perform decently, but they fail to incorporate volumetric information. Although 3D methods such as 3D UNet [7] and VoxResNet [6] are able to achieve satisfactory results in MRI PCa detection, 3D models inherently struggle against anisotropic data [15, 16]. In prostate mpMRI, the through-plane resolution (3–6 mm) is typically much poorer than the in-plane resolution (0.3–1.0 mm) [29], which poses challenges that undermine the ability of 3D models to perform accurate analysis. 2.5D methods, including CAT-Net [13], CSAM [14], and

AFTER-UNet [30], among others, were proposed to address these challenges. They primarily analyze images in 2D, while incorporating volumetric information through an attention mechanism, 3D convolution, concatenation, etc. This aligns with how radiologists analyze images to make clinical decisions, inasmuch as they mainly focus on an image of interest while also attending to and considering other slices in the volume. Nevertheless, these methods under-utilize the local information and have many hyperparameters to tune and/or parameters to train.

Despite the success of current deep learning-based approaches in MRI PCa detection in terms of accuracy and detection rate, there remains room for improvement. Measuring the uncertainty of predictions is crucial for enhancing the robustness of the models, potentially aiding radiologists in making more informed diagnoses and decisions, because it offers insights into how confident the network is about its per-pixel predictions. In particular, epistemic uncertainty is more important as it depicts model uncertainty [17]. Uncertainty in segmentation has been discussed in medical imaging [5, 28], but studies have focused on larger organs rather than much smaller cancer lesions. Moreover, most current research has employed techniques such as Monte Carlo (MC) dropout, ensemble, or an auxiliary branch to measure the epistemic uncertainty while relying on *softmax* for class probabilities even though this is known to inflate the probability of the predicted class due to the exponentiation of the outputs [25].

We address this problem from two perspectives. First, we aim to improve cross-slice attention by considering both global and local information. Second, we perform evidential deep learning (EDL) on PCa detection for improved epistemic uncertainty estimation. The main contributions of the work are as follows:

1. We devise a novel cross-slice attention mechanism, named Global-Local Cross-Slice Attention (GLCSA), leveraging both local and global information.
2. We introduce a new evidential critical (EC) loss for EDL for PCa detection, which to our knowledge is the first EDL method applied to the detection or segmentation of small objects.
3. We achieve state-of-the-art (SOTA) results across two different PCa datasets.
4. We show that our method for estimating epistemic uncertainty outperforms competing approaches, which can lead to improved detection results.

2 Preliminaries

2.1 2.5D Volumetric Image Analysis

We denote an input image volume $x_0 \in \mathbb{R}^{l \times c_0 \times h_0 \times w_0}$, where l is the total number of slices in the volume, c_0 is the number of channels, indicating the imaging modalities or types of imaging that are available, and h_0 and w_0 are the height and width of each slice. The pure 2D encoder E and decoder D in a standard encoder-decoder architecture for image segmentation are accompanied by three functions: $f_{\text{pre}}(\cdot)$ applied to the original volume, $f_{\text{mid}}(\cdot)$ applied to the feature maps between the encoder and decoder, and $f_{\text{post}}(\cdot)$ applied to the output of the decoder. 2.5D volumetric image segmentation is defined as follows:

For any E and D , given the segmentation output

$$y = f_{\text{post}}(D(f_{\text{mid}}(E(f_{\text{pre}}(x_0))))), \quad (1)$$

when at least one of $f_{\text{pre}}(\cdot)$, $f_{\text{mid}}(\cdot)$, and $f_{\text{post}}(\cdot)$ involves operations between different slices, the segmentation model is 2.5D [14].

Conventional 2.5D models stack nearby slices as input to segment the slice of interest [9]. In these methods $f_{\text{pre}}(\cdot)$ stacks nearby slices, while $f_{\text{mid}}(\cdot)$ and $f_{\text{post}}(\cdot)$ are identity functions. In contrast, SAUNet [31] performs attention across different slices in $f_{\text{post}}(\cdot)$, with $f_{\text{pre}}(\cdot)$ and $f_{\text{mid}}(\cdot)$ being identity functions. More recent works, such as CAT-Net [13], AFTER-UNet [30], CSAM [14], etc., conduct cross-slice attention between the encoder and decoder, where $f_{\text{mid}}(\cdot)$ is the cross-slice attention module, while $f_{\text{pre}}(\cdot)$ and $f_{\text{post}}(\cdot)$ are identity functions.

2.2 Evidential Deep Learning

Sensoy et al. [25] proposed EDL based on the Dempster-Shafer Theory (DST) of Evidence [26, 27] to explicitly model the epistemic uncertainty of each point prediction by assigning a set of belief masses $b_k \geq 0$ for any $k \in [1, K]$ representing a set of exclusive possible states; i.e., the possible class labels at a given pixel. The belief masses, associated with the Dirichlet distribution, along with the overall uncertainty mass $u \geq 0$ (the ‘‘I don’t know’’ class) should sum to 1:

$$u + \sum_{k=1}^K b_k = 1. \quad (2)$$

Belief b_k for singleton k and uncertainty u may be calculated as

$$b_k = \frac{e_k}{S}, \quad u = \frac{K}{S}, \quad S = K + \sum_{k=1}^K e_k, \quad (3)$$

where e_k is the evidence, which quantifies the level of support gathered from the data for classifying the sample into class k . Equation (3) demonstrates that increased evidence leads to decreased uncertainty, and in the absence of any evidence, $u = 1$, indicating complete uncertainty. This is associated with the Dirichlet distribution by $\alpha_k = e_k + 1$, where α_k is the Dirichlet concentration parameter for class k . The Dirichlet distribution is the conjugate prior of the categorical distribution, so the probability assignment of the categorical distribution over all possible classes can be sampled according to

$$\mathbf{p} \sim \text{Dir}(\mathbf{p}|\boldsymbol{\alpha}), \quad \hat{\mathbf{y}} \sim \text{Cat}(\mathbf{y}|\mathbf{p}), \quad (4)$$

where $\text{Dir}(\mathbf{p}|\boldsymbol{\alpha})$ is the Dirichlet distribution over probabilities $\mathbf{p}$ given concentration parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)$ and $\text{Cat}(\mathbf{y}|\mathbf{p})$ is the categorical distribution over classes $\mathbf{y}$ in one-hot encoding given the probabilities for each class $\mathbf{p}$. The

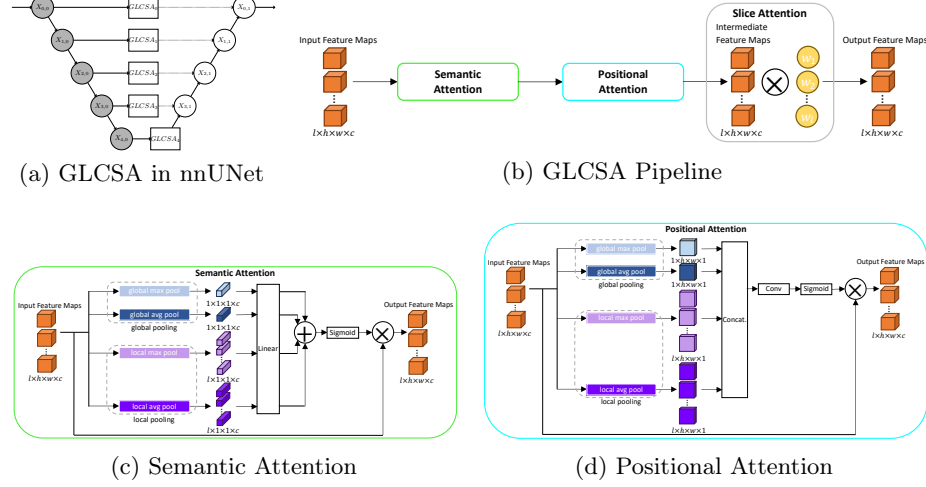


Fig. 1: GLCSA implementation.

expected probability of singleton k is the mean of the corresponding Dirichlet distribution:

$$\hat{p}_k = \frac{\alpha_k}{S}. \quad (5)$$

EDL networks replace the conventional *softmax* function in the final layer with an alternative activation function whose output is taken as the evidence vector $\mathbf{e} = (e_1, e_2, \dots, e_K)$, which is then used to calculate the concentration parameters $\boldsymbol{\alpha}$ that parameterize the Dirichlet distribution. Thus, standard deep learning directly predicts the class probability distribution, while EDL predicts the density of the class probability in the categorical distribution by modeling a Dirichlet distribution. As a result, EDL models the second-order probabilities, giving us more reasonable class probabilities and uncertainty estimation.

3 Global-Local Cross-Slice Attention

GLCSA has three parts: semantic, positional, and slice attention, organized sequentially. It can be inserted into any UNet-like architecture at the skip connection. 2.5D GLCSA processes the feature maps generated by the encoder and forwards its outputs to the decoder; i.e., GLCSA is $f_{\text{mid}}(\cdot)$ in (1). Fig. 1a illustrates the implementation of GLCSA in UNet-like architectures and Fig. 1b illustrates the GLCSA pipeline.

3.1 Semantic Attention

As shown in Fig. 1c, semantic attention processes input feature maps to re-weight them based on the relevance of semantic information across all the 2D feature

maps of the volume. The input feature maps have dimensions $l \times h \times w \times c$, where l is the number of slices in the volume, h and w are the height and width, and c is the number of channels. To perform semantic attention, the feature maps are first squeezed in the l , h , and w dimensions for global max and average pooling, and in the h and w dimensions for local max and average pooling. The outputs of the pooling layers are then fed into a linear layer and subsequently added together before going into the sigmoid function to produce the attention maps.

3.2 Positional Attention

Positional attention, illustrated in Fig. 1d, takes as input the output of semantic attention and adjusts the feature maps according to the significance of positions within each slice. The input features maps are first squeezed in the l and c dimensions for global max and average pooling, and in only the c dimension for local max and average pooling. They are then concatenated in the c dimension before passing a learnable convolutional kernel, whose output passes through a sigmoid function to produce the attention map that re-weights the input.

3.3 Slice Attention

Slice attention, illustrated in Fig. 1b, weights each slice based on the relative position of the slice. For instance, slices at the top and bottom of the volume, should receive lower weights as they are less likely to intersect the prostate, whereas central slices should be assigned higher weights as they are more likely to span it. This attention mechanism employs learnable weights w_i , where $1 \leq i \leq l$, each of which multiplies the feature maps $F_i \in \mathbb{R}^{h \times w \times c}$ output from the positional attention for slice i . The re-weighted feature maps are then fed into the decoder.

4 Evidential Critical Loss

EDL effectively quantifies epistemic uncertainty without extra computational cost; however, conventional EDL optimizes the Bayes risk of either cross-entropy loss or sum of squared loss [25]. These training objectives fail to account for the disproportionately small size of PCa lesions relative to the imaging volume. The number of background pixels outnumbers the lesion pixels by a factor of 1K to 10K in patients with suspicious lesions, making accurate PCa detection challenging and leading to potential overfitting to the background. Therefore, we introduce a new training objective, the EC loss, for evidential deep learning on highly imbalanced datasets. This loss function prioritizes critical classes, i.e., classes that are hard to classify, by reducing the loss impact from easier examples.

For a pixel i , we let $\mathbf{y}_i$ and $\mathbf{p}_i$ be the one-hot encoding vectors for the ground truth and prediction, respectively, and denote the concentration parameters as $\boldsymbol{\alpha}_i$. The conventional evidential loss can be written as

$$L_{ev,i} = \int \|\mathbf{y}_i - \mathbf{p}_i\|_2^2 \frac{\prod_{j=1}^K p_{ij}^{\alpha_{ij}-1}}{B(\boldsymbol{\alpha}_i)} d\mathbf{p}_i = \sum_{j=1}^K \left((y_{ij} - p_{ij})^2 + \frac{p_{ij}(1 - p_{ij})}{S_i + 1} \right), \quad (6)$$

where K is the number of classes, $B(\cdot)$ is the Beta function, and $S_i = \sum_{j=1}^K \alpha_{ij}$. However, this would result in the model overfitting the easy classes. Therefore, a modulation factor $(1 - \lambda_i)^\gamma$ and a weighting factor β_i are introduced to prioritize the critical classes [4, 18]; i.e., emphasizing the loss at true lesion pixels:

$$L_{\text{EC},i} = (1 - \lambda_i)^\gamma \beta_i L_{\text{ev},i}, \quad (7)$$

where $\lambda_i = \exp(\sum_{j=1}^K y_{ij} \log p_{ij})$ is the likelihood of the ground truths given the model predictions and γ controls the modulation intensity. When the likelihood λ_i is high, the pixel is well-classified and the EC loss is down-weighted close to 0, whereas a low likelihood λ_i suggests that the pixel is incorrectly classified, and the EC loss is less affected by the modulation. Thus, the EC loss targets critical pixels; i.e., those prone to incorrect classification. The modulation parameter γ controls the rate at which the non-critical pixels are down-weighted, with larger values reducing the effect of non-critical pixels on the loss. Weighting factor β_i controls which pixels the loss prioritizes, with higher values assigned to lesion pixels to increase the penalty for missed detection.

False evidence may arise during training, but the EC loss alone cannot eliminate it. Ideally, the total evidence should shrink to 0 if the model is not able to correctly classify the specific pixel. Therefore, an additional regularizer is needed, which incorporates a Kullback-Leibler (KL) divergence into the training objective, which penalizes evidence that does not contribute to the data fit. Finally, the total loss is formulated as

$$L = \sum_{i=1}^N L_{\text{EC},i} + s \sum_{i=1}^N KL(\text{Dir}(\mathbf{p}_i | \boldsymbol{\alpha}_i) || \text{Dir}(\mathbf{p}_i | \mathbf{1})), \quad (8)$$

where s is the annealing coefficient during training, $KL(\cdot || \cdot)$ is the KL divergence between two distributions, and $\mathbf{1}$ is a vector of ones.

5 Experiments

5.1 Datasets

We used two datasets, including our internal dataset (*Internal*) and the PICA challenge [24] dataset (*PICAI*). *Internal* comprises 652 patients with T2-weighted imaging (T2WI), apparent diffusion coefficient (ADC), high b -value (HBV) ($b = 1400$) diffusion-weighted imaging (DWI) MRI, along with manual zonal segmentation annotations. T2WI, ADC, HBV, and zonal segmentation are stacked as the model input. *PICAI* comprises 1,293 patients who underwent T2WI, ADC, HBV ($b > 1000$) DWI MRI, annotated by clinical experts. Only T2WI, ADC, and HBV are stacked as the model input.

5.2 Implementation Details

We implemented our method on nnUNet [15] (Fig. 1). We use the segmentation model to perform PCa detection. The local maxima in the output probability

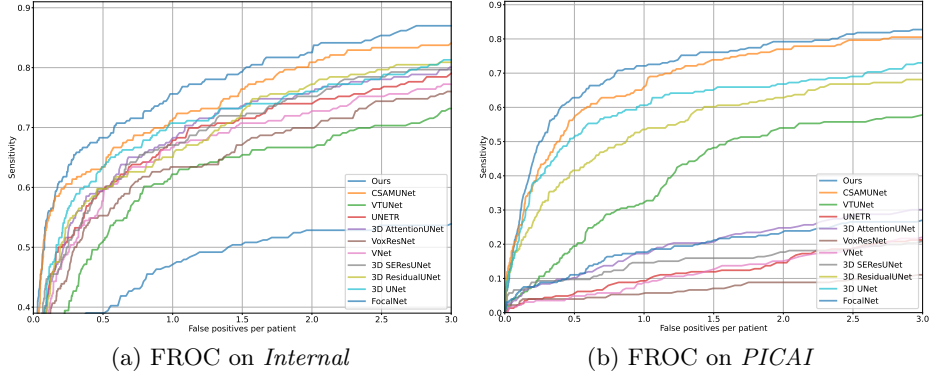


Fig. 2: Quantitative results.

map are considered the cancer detection points. The true positive PCa detection point is defined when it is within 5 mm of any PCa ground truth so as to account for a potential mismatch between the real lesion and the label due to labeling errors. For the EC loss, $\beta_i = 30$ if pixel i is a lesion pixel and $\beta_i = 1$ if it is a background pixel. The value for β_i is from the weighting in focal loss provided in AtPCa-Net [32]. On *Internal* $\gamma = 2.5$, while on *PICA* $\gamma = 4$. We utilized focal loss [18] for all the other experiments, and we set the focusing parameter to 2 for *Internal* and to 3 for *PICA*. We used the Adam optimizer with a learning rate of 0.0001 and weight decay regularization set to 1×10^{-5} . For all experiments, we used 5-fold cross-validation.

5.3 Quantitative Results

The Free-response Receiver Operating Characteristic (FROC) curves in Fig. 2 show that for both datasets our method outperforms all the competing methods; i.e., FocalNet [4], 3D UNet [7], 3D ResidualUNet [2], 3D SEResUNet [12], 3D AttentionUNet [20], VNet [19], VoxResNet [6], UNETR [10], and VTUNet [21]. The absence of zonal segmentation in the inputs in *PICA* hampers many competing methods [11], whereas our method consistently delivers the best results regardless of the availability of zonal segmentation. Our supplementary materials include qualitative results.

5.4 Ablation Study

Table 1 reports the results of our ablation study on *Internal*, comparing PCa detection sensitivities at different false positives (FPs)/patient. GCSA indicates global cross-slice attention; i.e., GLCSA without the local poolings. By introducing in turn GCSA, EC loss, and local poolings, the study confirmed that each component plays a role in improving performance.

Table 1: Ablation study on *Internal*.

FPS/Patient	0.5	1	1.5	2	2.5	3
nnUNet	0.659	0.715	0.740	0.768	0.789	0.801
nnUNet+GCSA	0.654	0.724	0.772	0.813	0.833	0.841
nnUNet+GCSA+EC loss	0.671	0.760	0.805	0.813	0.833	0.850
nnUNet+GLCSA+EC loss	0.687	0.760	0.801	0.841	0.854	0.870

Table 2: Performances of different uncertainty methods on *PICAI*.

FPS/Patient	0.5	1	1.5	2	2.5	3
MC Dropout	0.540	0.677	0.717	0.743	0.752	0.779
EDL	0.642	0.726	0.761	0.792	0.819	0.827

5.5 Uncertainty Estimation

Applying our proposed architecture on *PICAI*, Table 2 shows that our EDL-based method outperforms the MC dropout-based methods. The supplementary materials include visual results of different uncertainty estimations on *Internal*.

Furthermore, we evaluated the effectiveness of our EDL-based method. The pixels were first grouped into 10 bins based on their uncertainty values, where each bin has a 0.1 interval in uncertainty. Then, the classification accuracy for each bin was calculated. Ideally, the accuracy should drop as the uncertainty increases. The results on *Internal* are shown in Fig. 3a, where EDL achieves better uncertainty estimation since the accuracy drops as uncertainty increases, whereas uncertainty by MC dropout does not provide information regarding the classification result. Our method not only has better PCa detection performance, but also better uncertainty estimation than MC dropout.

Additionally, we investigated how uncertainty estimation can further help with PCa detection. We set thresholds such that any pixel with a higher uncertainty value is disregarded during PCa detection. We tested thresholds 0.7, 0.8, 0.9, and 1.0, which is equivalent to no threshold. The plots in Fig. 3b show that when using a lower threshold—i.e., eliminating more pixels—the sensitivity increases when the FPS/patient is relatively low. However, when the FPS/patient increases, the difference in sensitivity for different thresholds decreases. Thus, our uncertainty estimation can successfully tell where the model may be wrong, and that information can further improve detection sensitivity.

6 Conclusions

We introduced GLCSA, a novel cross-slice attention mechanism that utilizes both global and local information, as well as a new EC loss that enables EDL in PCa detection. Experimenting with two different datasets, we showed that our proposed method outperforms other SOTA methods in terms of FROC under

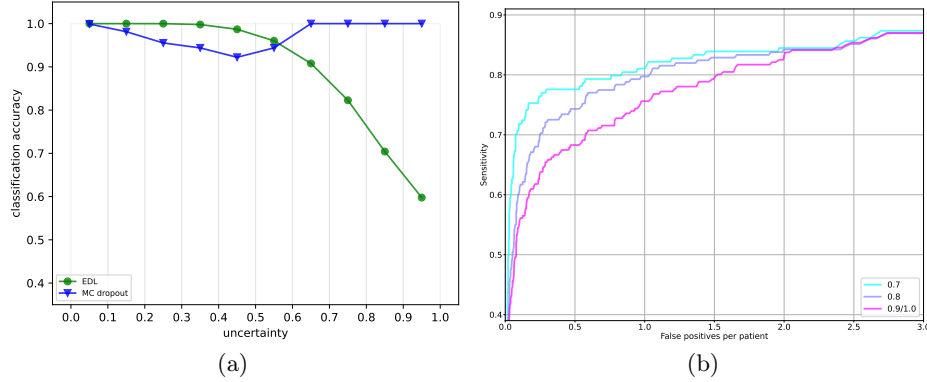


Fig. 3: (a) Comparison of EDL and MC dropout on uncertainty calibration, and (b) FROC curves for different uncertainty thresholds.

simple training settings and data augmentation. Our ablation study demonstrated that all key components contributed to improving performance. The uncertainty estimation inherent to our method is superior to its counterparts as it can improve the explainability and interpretability of the model while improving its performance when some pixels are filtered out by thresholding uncertainty values.

Acknowledgments. The research reported herein was funded in part by the National Institutes of Health under grants R01-CA248506 and R01-CA272702 and by the Integrated Diagnostics Program of the Departments of Radiological Sciences and Pathology in the UCLA David Geffen School of Medicine.

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

References

1. Appayya, M.B., Adshead, J., Ahmed, H.U., Allen, C., Bainbridge, A., Barrett, T., Giganti, F., Graham, J., Haslam, P., Johnston, E.W., et al.: National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection—recommendations from a UK consensus meeting. *BJU International* **122**(1), 13 (2018)
2. Bhalerao, M., Thakur, S.: Brain tumor segmentation based on 3D residual U-Net. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 218–225 (2020)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation. In: *European conference on computer vision*. pp. 205–218. Springer (2022)
4. Cao, R., Bajgirani, A.M., Mirak, S.A., Shakeri, S., Zhong, X., Enzmann, D., Raman, S., Sung, K.: Joint prostate cancer detection and gleason score prediction in mp-MRI via FocalNet. *IEEE Transactions on Medical Imaging* **38**(11), 2496–2506 (2019)

5. Carannante, G., Dera, D., Bouaynaya, N.C., Rasool, G., Fathallah-Shaykh, H.M.: Trustworthy medical segmentation with uncertainty estimation. arXiv preprint arXiv:2111.05978 (2021)
6. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* **170**, 446–455 (2018)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432. Springer (2016)
8. Duran, A., Dussert, G., Rouvière, O., Jaouen, T., Jodoin, P.M., Lartizien, C.: ProstAttention-Net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. *Medical Image Analysis* **77**, 102347 (2022)
9. Han, L., Chen, Y., Li, J., Zhong, B., Lei, Y., Sun, M.: Liver segmentation with 2.5D perpendicular UNets. *Computers & Electrical Engineering* **91**, 107118 (2021)
10. 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)
11. Hosseinzadeh, M., Saha, A., Brand, P., Slootweg, I., de Rooij, M., Huisman, H.: Deep learning-assisted prostate cancer detection on bi-parametric MRI: Minimum training data size requirements and effect of prior knowledge. *European Radiology* pp. 1–11 (2022)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR. pp. 7132–7141 (2018). <https://doi.org/10.1109/CVPR.2018.00745>
13. Hung, A.L.Y., Zheng, H., Miao, Q., Raman, S.S., Terzopoulos, D., Sung, K.: CAT-Net: A cross-slice attention transformer model for prostate zonal segmentation in MRI. *IEEE Transactions on Medical Imaging* **42**(1), 291–303 (2022)
14. Hung, A.L.Y., Zheng, H., Zhao, K., Du, X., Pang, K., Miao, Q., Raman, S.S., Terzopoulos, D., Sung, K.: CSAM: A 2.5D cross-slice attention module for anisotropic volumetric medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5923–5932 (2024)
15. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnU-Net: Self-adapting framework for U-Net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
16. Jia, H., Xia, Y., Song, Y., Zhang, D., Huang, H., Zhang, Y., Cai, W.: 3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images. *IEEE Transactions on Medical Imaging* **39**(2), 447–457 (2019)
17. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988 (2017)
19. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
20. 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)

21. Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M.: A robust volumetric transformer for accurate 3D tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 162–172. Springer (2022)
22. Rawla, P.: Epidemiology of prostate cancer. *World Journal of Oncology* **10**(2), 63 (2019)
23. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
24. Saha, A., Twilt, J.J., Bosma, J.S., van Ginneken, B., Yakar, D., Elschot, M., Veltman, J., Fütterer, J., de Rooij, M., Huisman, H.: Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol) (2022). <https://doi.org/10.5281/zenodo.6667655>
25. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. *Advances in Neural Information Processing Systems* **31** (2018)
26. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton University Press (1976)
27. Shafer, G.: Dempster-Shafer theory. *Encyclopedia of Artificial Intelligence* **1**, 330–331 (1992)
28. Tang, P., Yang, P., Nie, D., Wu, X., Zhou, J., Wang, Y.: Unified medical image segmentation by learning from uncertainty in an end-to-end manner. *Knowledge-Based Systems* **241**, 108215 (2022)
29. Turkbey, B., Rosenkrantz, A.B., Haider, M.A., Padhani, A.R., Villeirs, G., Macura, K.J., Tempany, C.M., Choyke, P.L., Cornud, F., Margolis, D.J., et al.: Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *European Urology* **76**(3), 340–351 (2019)
30. Yan, X., Tang, H., Sun, S., Ma, H., Kong, D., Xie, X.: AFter-UNet: Axial fusion transformer UNet for medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3971–3981 (2022)
31. Zhang, Y., Yuan, L., Wang, Y., Zhang, J.: SAU-Net: efficient 3D spine MRI segmentation using inter-slice attention. In: Medical Imaging With Deep Learning. pp. 903–913. PMLR (2020)
32. Zheng, H., Hung, A.L.Y., Miao, Q., Song, W., Scalzo, F., Raman, S.S., Zhao, K., Sung, K.: AtPCa-Net: Anatomical-aware prostate cancer detection network on multi-parametric MRI. *Scientific Reports* **14**(1), 5740 (2024)