FAU-Net: An Attention U-Net Extension with Feature Pyramid Attention for Prostate Cancer Segmentation

Pablo Cesar Quihui-Rubio[1], Daniel Flores-Araiza[1], Miguel Gonzalez-Mendoza[1], Christian Mata[2,3], and Gilberto Ochoa-Ruiz[1](✉)

[1] School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, Mexico
gilberto.ochoa@tec.mx
[2] Universitat Politècnica de Catalunya, 08019 Barcelona, Catalonia, Spain
[3] Pediatric Computational Imaging Research Group, Hospital Sant Joan de Déu, 08950 Esplugues de Llobregat, Catalonia, Spain

Abstract. This contribution presents a deep learning method for the segmentation of prostate zones in MRI images based on U-Net using additive and feature pyramid attention modules, which can improve the workflow of prostate cancer detection and diagnosis. The proposed model is compared to seven different U-Net-based architectures. The automatic segmentation performance of each model of the central zone (CZ), peripheral zone (PZ), transition zone (TZ) and Tumor were evaluated using Dice Score (DSC), and the Intersection over Union (IoU) metrics. The proposed alternative achieved a mean DSC of 84.15% and IoU of 76.9% in the test set, outperforming most of the studied models in this work except from R2U-Net and attention R2U-Net architectures.

Keywords: Segmentation · U-Net · Attention · Uncertainty Quantification · Prostate Cancer · Deep Learning

1 Introduction

Prostate cancer (PCa) is the most common solid non-cutaneous cancer in men and is among the most common causes of cancer-related deaths in 13 regions of the world [9].

When detected in early stages, the survival rate for regional PCa is almost 100%. In contrast, the survival rate when the cancer is spread to other parts of the body is of only 30% [3]. Magnetic Resonance Imaging (MRI) is the most widely available non-invasive and sensitive tool for detection of PCa, due to its high resolution, excellent spontaneous contrast of soft tissues, and the possibility of multi-planar and multi-parametric scanning [5]. Although MRI is used traditionally for staging PCa, it can be also be used for the PCa detection through the segmentation of Regions of Interest (ROI) from MR images.

The use of image segmentation for PCa detection and characterization can help determine the localization and the volume of the cancerous tissue [7]. This highlights the importance of an accurate and consistent segmentation when detecting PCa.

H. Calvo et al. (Eds.): MICAI 2023, LNAI 14392, pp. 165–176, 2024.
https://doi.org/10.1007/978-3-031-47640-2_14

However, the most common and preferred method for identifying and delimiting prostate gland and prostate regions of interest is by performing a manual inspection by radiologists [1]. This manual process is time-consuming, and is sensitive to specialists' experience, resulting in a significant intra- and inter-specialist variability [14]. Automating this process for the segmentation of prostate gland and regions of interest, in addition to saving time for radiologists, can be used as a learning tool for others and have consistency in contouring [11].

Deep Learning (DL) base methods have recently been developed to perform automatic prostate segmentation [6]. One of the most popular methods is U-Net [16], which has been the inspiration behind many recent works in literature.

In this work, we propose an automatic prostate zone segmentation method that is based on an extension of Attention U-Net that combines two types of attention, pyramidal and additive. We also include the pixel-wise estimation of the uncertainty.

The zones evaluated in this work are the central zone (CZ), the peripheral zone (PZ), transition zone (TZ), and, in the case of a disease, the tumor (TUM), different from other works, which only evaluate CZ and PZ [10].

The rest of this paper is organized as follows: Sect. 2 describes previous works dealing with the prostate segmentation. Section 3 describes the dataset used in this work, the proposed architecture, as well as the experimental setup to evaluate it. In Sect. 4 the results of the experiments are presented and discussed and Sect. 5 concludes the article.

2 State-of-the-Art

In medical imaging, one of the best known DL models in the literature for segmentation is U-Net, which consists of two sub-networks: an encoder with a series of four convolutions and max-pooling operations to reduce the dimension of the input image and to capture its semantic information at different levels. The second sub-network is a decoder that consists of four convolution and up-sampling operations to recover the spatial information of the image [16]. The work from Zhu et *al.* [18] proposes a U-Net based network to segment the whole prostate gland, obtaining encouraging results. Moreover, this architecture has served as the inspiration for some variants that enhance the performance of the original model. One example is the work from Oktay et *al.* [13], which proposes the addition of attention gates inside the original U-Net model with the intention of making the model focus on the specific target structures. In this architecture, the attention layers highlight the features from the skip connections between the encoder and the decoder. Many others extension architectures have been proposed since U-Net was released, some of them include Dense blocks [17], residual and recurrent blocks [2], even novel architectures implemented transformers blocks named Swin blocks in order to obtain Swin U-Net [4].

All the mentioned models had demonstrated great results in many biomedical image datasets. However, in this work we focused on PCa segmentation, in particular, the main zones of the prostate, which has not been deeply investigated by some of these models.

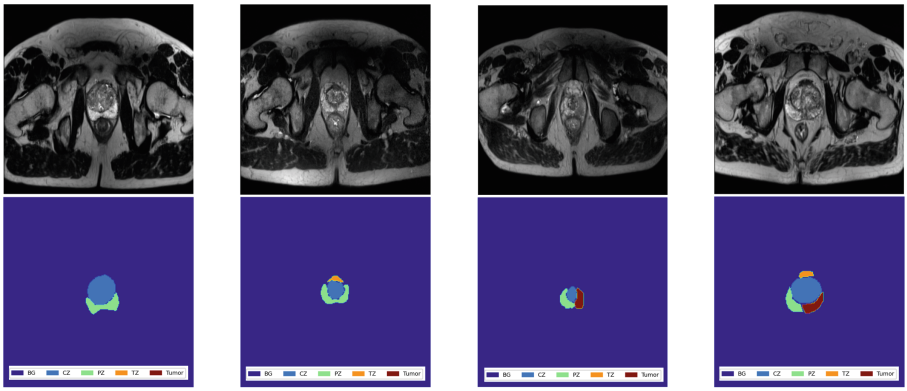
3 Materials and Methods

3.1 Dataset

This study was carried out in compliance with the Centre Hospitalier de Dijon. The dataset provided by these institutions consists of three-dimensional T2-weighted fast spin-echo (TR/TE/ETL: 3600 ms/ 143 ms/109, slice thickness:1.25 mm) images acquired with sub-millimetric pixel resolution in an oblique axial plane. The total number of patients from the dataset are 19, with a total of 205 images with their corresponding masks used as a ground truth. The manual segmentation of each with four regions of interest (CZ, PZ, TZ, and TUMOR) was also provided, this process was cautiously validated by multiple professional radiologists and experts using a dedicated software tool [12, 15].

The entire dataset contains four different combination of zones, being: (CZ+PZ), (CZ+PZ+TZ), (CZ+PZ+Tumor), and (CZ+PZ+TZ+Tumor) with 73, 68, 23, and 41 images respectively. For the purpose of this work, the dataset was divided in 85% for training and 15% for testing, keeping a similar distribution in both sets of data, having a total of 174 images for training, and 31 for testing.

In Fig. 1 examples of images from every possible combination of zones in the dataset are presented.



(a) CZ and PZ (b) CZ, PZ, and TZ (c) CZ, PZ, and Tumor (d) CZ, PZ, TZ, and Tumor

Fig. 1. Sample images from every possible combination of zones in the dataset are presented in the upper row. Their respective ground truth masks are shown in the lower row.

3.2 Feature Pyramid Attention

The work of Yonkai et *al.* [9] introduces the feature pyramid attention (FPA) network to capture information at multiple scales. It contains three convolutional blocks of different

sizes (3×3, 5×5 and 7×7) to extract the features from different scales. These are then integrated from smaller to bigger, to incorporate the different scales. In our work, the attention map is multiplied by the features from the skip connection after a 1×1 convolution. A visual representation of this attention block is presented in Fig. 2.

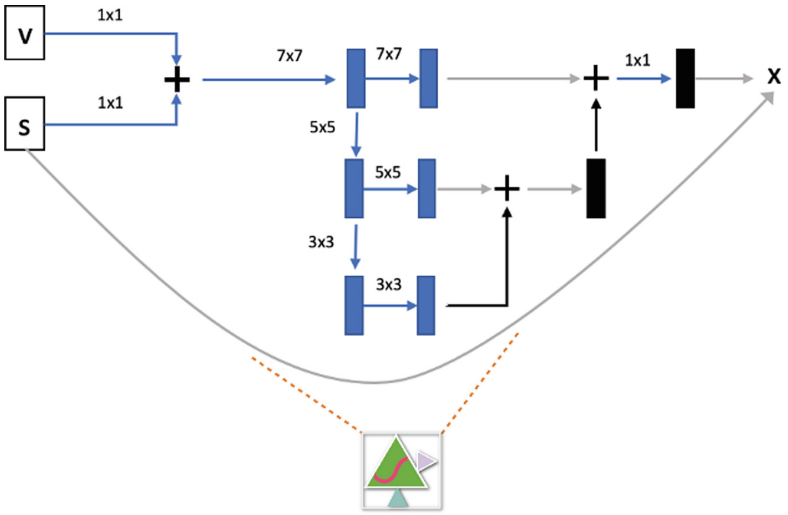


Fig. 2. The feature pyramid attention block. It consists of three convolutional blocks of 3×3, 5×5, and 7×7 which responses are integrated to capture the context of each level.

3.3 Proposed Work

This contribution proposes the Fusion Attention U-Net (FAU-Net), an Attention-U-Net-based extension with pyramidal and additive attention. The proposed model is used to perform the segmentation of five different regions from the PCa dataset described in Sect. 3.1.

Attention U-Net implements attention gates (AG) into the U-Net architecture to highlight salient features that are passed through the skip connections, these gates allow the network to disambiguate irrelevant and noisy responses in skip connections, leaving only the relevant activations to merge [13]. In the architecture proposed, we used AGs in the last three levels of the architecture. Meanwhile, in the first level, the implementation of a FPA was carried out to give further attention in those layers, were more data could be leaked. In Fig. 3 an entire representation of the architecture is shown.

A comparison between U-Net [16], Attention U-Net [13], Dense U-Net [17], Attention Dense U-Net [8], R2U-Net [2], Attention R2U-Net, Swin U-Net [4] and the proposed FAU-Net was done to validate the results obtained.

Most of the works in the literature perform the segmentation task of only two zones, and the number of works that consider a third zone (TZ) is limited, mainly because the

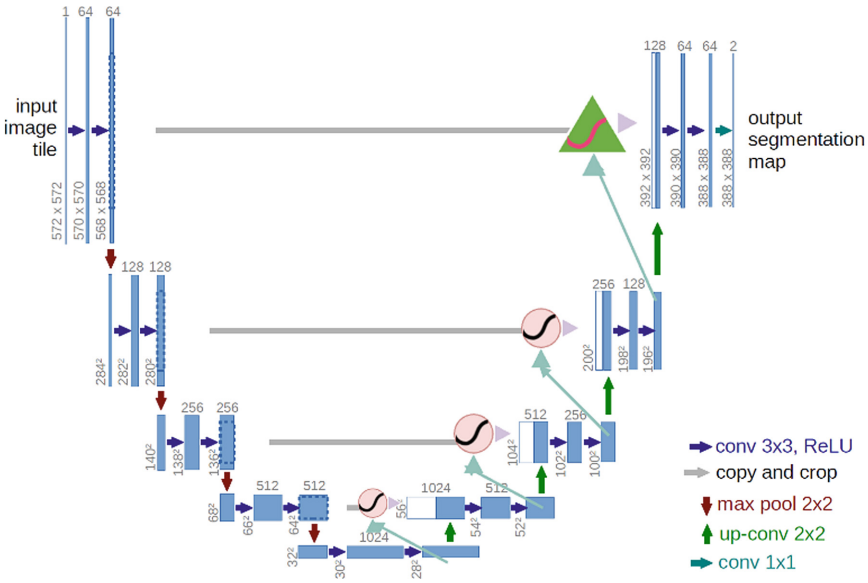


Fig. 3. Proposed Fusion Attention U-Net model. The input image first goes through the contracting path. The boxes represent the feature map at each layer, and the blue boxes represent the cropped feature maps from the contracting path. (Color figure online)

boundaries are more delimited than zones such as TZ or Tumor. In this work we used a private dataset which incorporates the TZ and, in some cases, a tumor. This zone is important because it could lead to a proper diagnosis or treatment if a tumor is present.

Therefore, we proposed an attention-based model to perform segmentation with a dataset of only T2-weighted images with 4 prostate zones, and compare the results against other models proposed in the literature. We analyzed the segmentation of the prostate zones using different metrics to choose the best DL architecture. Finally, we did a qualitative analysis of the predictions of each model.

In Table 1 is shown the number of parameters, which are different for each model, being the one with the lowest number the original U-Net, and the Swin U-Net with the highest number of parameters. FAU-Net has only around 160,000 more parameters than U-Net and Attention U-Net, being the third model with less parameters.

All the models were trained with the same dataset for 145 epochs, using Adam optimizer with a learning rate of 0.0001, batch size of 6, and categorical cross-entropy as loss function. The performance was evaluated using F1-score and Intersection over Union (IoU) as the main metrics. All the training was done using a NVIDIA DGX workstation, using a V100 GPU.

Table 1. Count of trainable parameters for each model analyzed during this work.

Model	Number of parameters
U-Net	1,940,885
Attention U-Net	1,995,409
FAU-Net	**2,158,505**
Dense U-Net	4,238,389
Attention Dense U-Net	4,271,521
R2U-Net	6,003,077
Attention R2U-Net	6,036,081
Swin U-Net	26,598,344

4 Results and Discussion

The results of this work are divided in two subsections for further analysis and comparison between the models: quantitative and qualitative.

4.1 Quantitative Results

Table 2 shows a summary of results for the evaluation of the eight studied architectures in two metrics (DSC and IoU) and loss value. Each evaluation corresponds to the mean value of the metrics for all the prostate zones and images in the test set. The bold values represent the model that achieved the best metric score within all of them.

Table 2. The model performance evaluation was conducted using the Categorical Cross-Entropy (CCE) as the loss function. The metrics were designated with either an upward (↑) or downward (↓) arrow to indicate whether higher or lower values were desirable. Bold values and green highlights denote the best metric score achieved among all models.

Model	IoU ↑	DSC ↑	Loss ↓
U-Net	70.76	80.00	0.0138
Dense U-Net	74.53	83.65	0.0225
Swin U-Net	75.24	83.91	0.0124
Attention U-Net	74.92	84.01	0.0114
Attention Dense U-Net	75.12	84.01	0.0211
FAU-Net	75.49	84.15	**0.0107**
R2U-Net	76.60	85.30	0.0131
Attention R2U-Net	**76.89**	**85.42**	0.0120

As expected, the extended U-Net architectures performed better than the original U-Net architecture. For instance, the Dense U-Net model showed an improvement of

approximately 5% in both metrics. However, the Swin U-Net model, based on Swin Transformers and considered one of the best architectures available, did not perform as well on the dataset used in this study. It outperformed U-Net and Dense U-Net models in both metrics by 6%, and Attention U-Net and Attention Dense U-Net in the IoU metric by only 0.4% and 0.1%, respectively. The subpar performance of this model could be attributed to various factors, but the most likely explanation is the small size of the dataset and the high number of training parameters, which may have led to overfitting.

Incorporating attention modules into U-Net and Dense U-Net models resulted in significant improvements compared to models without them. Attention U-Net outperformed U-Net by more than 5% in both metrics. Meanwhile, Attention Dense U-Net achieved the same DSC score as Attention U-Net and a higher IoU score by approximately 1%. These results indicate that attention modules are beneficial for obtaining better prostate segmentation, even with a relatively small dataset.

The proposed FAU-Net architecture in this study incorporated two types of attention: additive attention, as used in previous models, and pyramidal attention, consisting of attention modules in a cascade fashion. The objective of this model was to focus on the most complex features of each prostate image and obtain better information, and the results support this hypothesis. FAU-Net achieved IoU and DSC values of 75.49% and 84.15%, respectively, improving U-Net results by more than 6%. However, this architecture was surpassed by R2U-Net and Attention R2U-Net.

R2U-Net and Attention R2U-Net are architectures that rely on recurrent residual blocks, which aid in extracting more information from deeper image features. In this study, Attention R2U-Net was the top-performing model overall, achieving metric scores greater than 76% for IoU and 85% for DSC, with a loss value of 0.0120.

To gain a comprehensive understanding of the segmentation metrics in biomedical images, particularly related to the prostate, it is important to examine specific tissue zones. After analyzing the segmentation metrics through the full test set from the dataset, Fig. 4 shows the IoU scores obtained from each image in each prostate zone. Each model is represented by a different color, and each test image is represented by a colored dot with the corresponding value. However, it's essential to note that not all images in the set had the same distribution, resulting in fewer dots in the boxplot for prostate zones such as TZ and Tumor. Nonetheless, the performance trends of the models in each particular zone can be analyzed.

Undoubtedly, the central and peripheral zones are the easiest for all models to segment, with only a few images having low IoU values. However, segmenting the peripheral zone appears slightly more challenging, likely due to its smaller size. The proposed FAU-Net was the best model overall for these two zones, with a mean IoU score of 82.63% and 72.55% for CZ and PZ, respectively. In contrast, the worst model was U-Net, with values below 80% for CZ and 67% for PZ.

As for the transition zone and tumors, the variation between the models is more noticeable in Fig. 4. Most models had lower values for outliers in the transition zone, achieving mean IoU scores lower than 60% in all of them except R2U-Net, which managed to reach a mean score of 61% in TZ.

Prostate tumors are a challenging task for segmentation due to the different types of geometry and boundaries between other tissues or zones. However, unlike TZ, most of

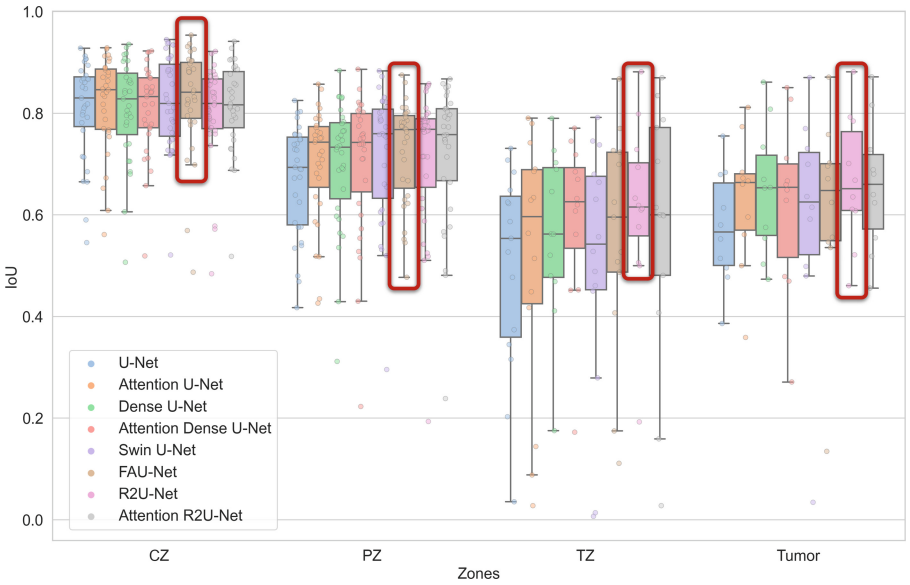


Fig. 4. The IoU scores obtained for each prostate zone from all images in the test set were compared between models. A line represents the median uncertainty value obtained, dots represent the particular score for each image, and the best model for each zone is indicated with a red box. (Color figure online)

the models managed not to have many outliers when segmenting the tumor, and most reached values higher than 60%. The worst model for segmenting the tumor was U-Net, with a mean IoU score of only 57%. On the other hand, the best model, R2U-Net, surpassed this model by 10%, obtaining a mean IoU score of 67%.

4.2 Qualitative Results

A visual inspection was carried out of the segmentation results of the eight models discussed in this study. This analysis of results complements the previous quantitative analysis based on the metrics. In this inspection, the images from the test set were visually compared to their corresponding ground truth, and conclusions were stated.

Figure 5 presents a qualitative comparison between each model's prediction in four different example images from the dataset, with all the possible combinations of zones. The first two rows show the original T2-MRI image of the prostate and below its corresponding ground truth. Then, each row represents prediction of the different models.

Starting from the base model U-Net, it is clear that U-Net had difficulty correctly segmenting all pixels, especially in images with tumors, for example in image C, this model missed many pixels that corresponded to the Tumor; this could be a wrong lead for a radiologist who is relying upon this model. Nevertheless, even though a Tumor is present in example D, U-Net segments most of the pixels better than in the previous example, at least from a visual perspective.

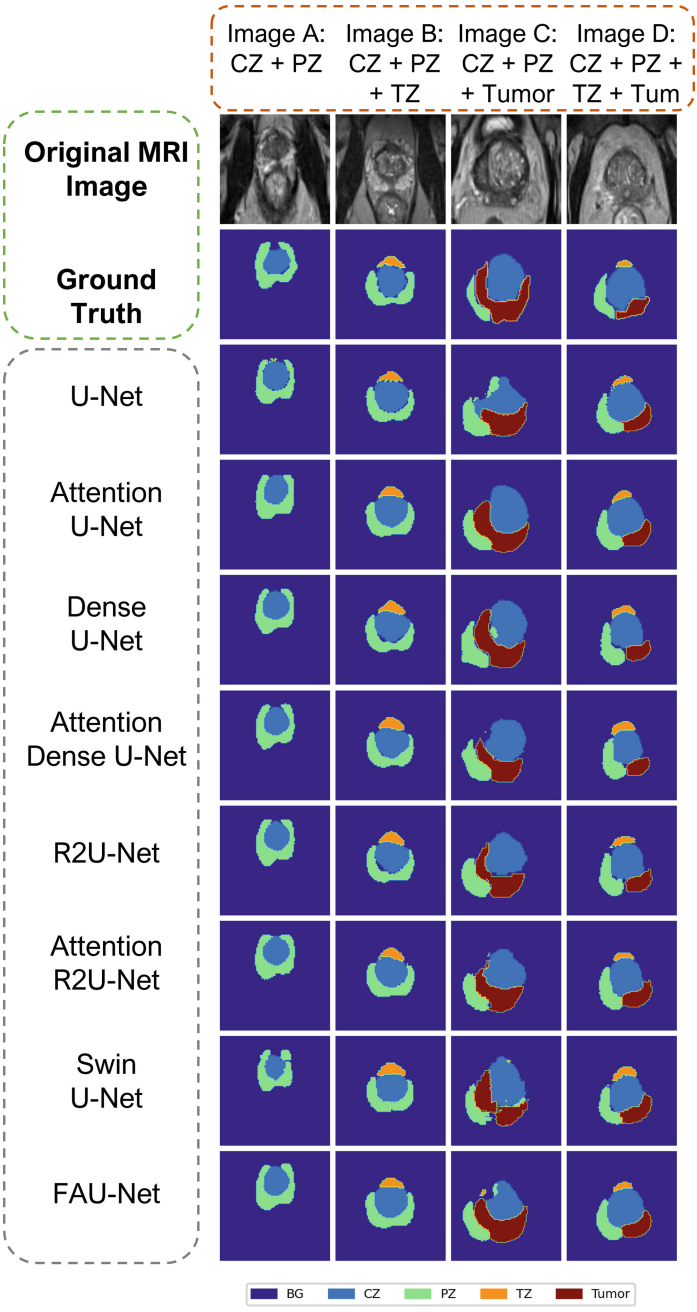


Fig. 5. Image comparison of segmentation results using U-Net-like architectures. All possible combinations of zones available in the dataset are used as examples for conducting predictions on MRI images.

Based on qualitative analysis, some models, such as Attention U-Net, R2U-Net, and FAU-Net, performed better in segmenting all prostate zones, including the Tumor. Compared to the other models, these models produced smoother and more complete segmentation in images with three or more zones. However, it should be noted that FAU-Net misclassified some pixels as TZ in example C, which does not include TZ.

It is clear that images with only two zones (CZ and PZ) are easier to segment for all the models, which are the bigger and more present ones in the dataset. Some models in examples C and D include more pixels in the smaller zones, resulting in a smoother segmentation; although this looks great from visual analysis, compared to the ground truth, that prediction is incorrect; thus, relying solely on visual analysis is not advisable.

As a qualitative conclusion of the predictions based on the examples from Fig. 5, it can be demonstrated that Attention U-Net and R2U-Net are the models with the best segmentation performance overall. However, based on the metrics and a visual analysis from the entire test set, in general the best performance was obtained by FAU-Net, R2U-Net, and Attention R2U-Net.

5 Conclusion

In this work, we proposed a U-Net extension using two attention blocks: additive and pyramidal. From the results shown in Sect. 4, we can conclude that the proposed architecture, FAU-Net, outperforms most of the studied architectures in this work. Moreover, other alternatives like R2U-Net and Attention R2U-Net, are still better suited to perform over this particular dataset than the proposed architecture. Furthermore, FAU-Net presents great metrics score and although it struggles in particular zones like TZ and Tumor, it is the best model to segment the CZ and PZ regarding the segmentation metrics in our dataset.

Considering that the results obtained are promising, further investigation can be done by improving the FAU-Net architecture to achieve even better results. For instance, a future implementation of feature pyramid attention module in the R2U-Net architecture can lead to promising results using the dataset studied in this work for prostate segmentation. Also, trying more combinations of the attention modules and/or adding more levels to the architecture can produce interesting results.

Acknowledgments. The authors wish to acknowledge the Mexican Council for Science and Technology (CONAHCYT) for the support in terms of postgraduate scholarships in this project, and the Data Science Hub at Tecnologico de Monterrey for their support on this project.

This work has been supported by Azure Sponsorship credits granted by Microsoft's AI for Good Research Lab through the AI for Health program. The project was also supported by the French-Mexican ANUIES CONAHCYT Ecos Nord grant 322537.

References

1. 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. **10** (2020). https://doi.org/10.1038/s41598-020-71080-0

2. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation. CoRR abs/1802.06955 (2018)
3. AstraZeneca: A personalized approach in prostate cancer (2020). https://www.astrazeneca.com/our-therapy-areas/oncology/prostate-cancer.html. Accessed 17 Oct 2021
4. Cao, H., et al.: Swin-Unet: Unet-like pure transformer for medical image segmentation. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) ECCV 2022. LNCS, pp. 205–218. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-25066-8_9
5. Chen, M., et al.: Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined. Acta Radiologica **49**(5), 602–610 (2008). https://doi.org/10.1080/02841850802004983. https://www.tandfonline.com/doi/abs/10.1080/02841850802004983
6. Elguindi, S., et al.: Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. Phys. Imaging Radiation Oncol. **12**, 80–86 (2019). https://doi.org/10.1016/j.phro.2019.11.006. https://www.sciencedirect.com/science/article/pii/S2405631619300569
7. Haralick, R., Shapiro, L.: Image segmentation techniques. Comput. Vis. Graph. Image Process. **29**(1), 100–132 (1985). https://doi.org/10.1016/S0734-189X(85)90153-7. https://www.sciencedirect.com/science/article/pii/S0734189X85901537
8. Li, S., Dong, M., Du, G., Mu, X.: Attention Dense-U-Net for automatic breast mass segmentation in digital mammogram. IEEE Access **7**, 59037–59047 (2019). https://doi.org/10.1109/ACCESS.2019.2914873
9. Liu, Y., et al.: Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. IEEE Access **7**, 163626–163632 (2019). https://doi.org/10.1109/ACCESS.2019.2952534
10. Liu, Y., et al.: Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. IEEE Access **8**, 151817–151828 (2020). https://doi.org/10.1109/ACCESS.2020.3017168
11. Mahapatra, D., Buhmann, J.M.: Prostate MRI segmentation using learned semantic knowledge and graph cuts. IEEE Trans. Biomed. Eng. **61**(3), 756–764 (2014). https://doi.org/10.1109/TBME.2013.2289306
12. Mata, C., Munuera, J., Lalande, A., Ochoa-Ruiz, G., Benitez, R.: MedicalSeg: a medical GUI application for image segmentation management. Algorithms **15**(06) (2022). https://doi.org/10.3390/a15060200. https://www.mdpi.com/1999-4893/15/6/200
13. Oktay, O., et al.: Attention U-net: learning where to look for the pancreas (2018). https://doi.org/10.48550/ARXIV.1804.03999
14. Rasch, C., et al.: Human-computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. J. Digit. Imaging **24**(5), 794–803 (2011). https://doi.org/10.1007/s10278-010-9341-2
15. Rodríguez, J., Ochoa-Ruiz, G., Mata, C.: A prostate MRI segmentation tool based on active contour models using a gradient vector flow. Appl. Sci. **10**(18) (2020). https://doi.org/10.3390/app10186163. https://www.mdpi.com/2076-3417/10/18/6163
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

17. Wu, Y., Wu, J., Jin, S., Cao, L., Jin, G.: Dense-U-Net: dense encoder-decoder network for holographic imaging of 3D particle fields. Opt. Commun. **493**, 126970 (2021). https://doi.org/10.1016/j.optcom.2021.126970. https://www.sciencedirect.com/science/article/pii/S0030401821002200
18. Zhu, Q., Du, B., Turkbey, B., Choyke, P.L., Yan, P.: Deeply-supervised CNN for prostate segmentation. CoRR abs/1703.07523 (2017). http://arxiv.org/abs/1703.07523