Fine-tuned feature selection to improve prostate segmentation via a fully connected meta-learner architecture

1st Dimitris Zaridis
Dept. of Biomedical Research FORTH-BRI
Ioannina, Greece
dimzaridis@gmail.com

2st Eugenia Mylona
Dept. of Biomedical Research FORTH-BRI
Ioannina, Greece
mylona.eugenia@gmail.com

3st Nikolaos Tachos
Dept. of Biomedical Research FORTH-BRI
Ioannina, Greece
ntachos@gmail.com

4st Kostas Marias
Inst. of Computer Science and Electrical Computer Engineering FORTH
Heraklion, Greece
kmarias@ics.forth.gr

5st Manolis Tsiknakis
Inst. of Computer Science & Dept. of Electrical Computer Engineering FORTH and Hellenic Mediterranean University
Heraklion, Greece
tsiknaki@ics.forth.gr

6st Dimitios I. Fotiadis
Inst. of Biomedical Research & Unit of Medical Technology and Intelligent Information Systems FORTH-BRI & Materials Science and Engineering Dept., University of Ioannina
Ioannina, Greece
fotiadis@uoi.gr

2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-8791-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BHI56158.2022.9926929

***Abstract*—Precise delineation of the prostate gland on MRI is the cornerstone for accurate prostate cancer diagnosis, detection, characterization and treatment. The present work proposes a meta-learner deep learning (DL) network that combines the complexity of 3 well-established DL models and fine tune them in order to improve the segmentation of the prostate compared to the base learners. The backbone of the meta-learner consist the original U-net, Dense2U-net and Bridged U-net models. A model was added on top of the three base networks that has four convolutions with different receptor fields. The meta-learner outperformed the base-learners in 4 out of 5 performance metrics. The median Dice Score for the meta-learner was 89% while for the second best model it was 83%. Except for Hausdorff distance, where the meta-learner and Dense2U-net performed equally well, the improvement achieved in terms of average sensitivity, balanced accuracy, dice score and rand error, compared to the best performing base-learner, was 6%, 3%, 5% and 4%, respectively.**

***Index Terms*—Prostate Segmentation, Deep Learning, Ensembling, Fine Tuning**

I. Introduction

One of the most prevalent cancer type among men worldwide is the prostate cancer. Although the important decrease in mortality over the last decades, prostate cancer remains a leading cause of cancer-related death among men [1].

Precise delineation of the prostate gland on MRI is the cornerstone for accurate prostate cancer diagnosis, detection, characterization and subsequent treatment. However, this remains a challenging task due to the position and morphology of the prostate [2]. In fact, the small size of the prostatic gland, the vague boundaries of the organ, and the closeness to neighboring organs, are likely to obscure the performance of both humans and machines in segmenting this region. Manual labeling of the prostate is a daunting and labor-intensive task. It is well-known that not only it requires certain clinical experience but also often it suffers from considerable inter-observer variability with respect to the segmented region [3].

With the breakthrough of Deep Learning (DL) in medical image segmentation, several algorithms based on Convolutional Neural Networks (CNN) have been developed to automatically and accurately segment the prostate. The U-net encoder-decoder architecture [4], in particular, has emerged at the standard DL architecture for image segmentation tasks and the backbone for a plethora of other emerging networks. In general, these models use a variable number of hidden layers between the input and output, and the nodes are connected in a serial and a parallel way to one another with varying weights in order to extract several redundant and low-level features at different steps and different scales. Although, to date, DL networks have demonstrated sufficient performance for prostate segmentation, their generalizability on external datasets remains an open question. Furthermore, these networks frequently fail to map the long-range relationships of local features, resulting in discriminative feature mappings in the resultant segmented picture for each class.

In the present work we propose a meta-learner DL network that combines the complexity of 3 well established models.Task specific performance is increased due to the fine tuning and important features alleviation by each of them.

Authorized licensed use limited to: Foundation for Research and Technology (FORTH). Downloaded on June 15,2023 at 08:40:01 UTC from IEEE Xplore. Restrictions apply.

II. METHODOLOGY

A. Imaging Datasets

In this work two publicly available datasets were used for model training, validation and testing. Specifically, for model training and validation, 204 patients from the Prostate-X2 dataset along with the corresponding masks of the whole gland were utilized [5]. Prostate-X dataset consist of 3206 frames in T2-Weighted MR images from Siemens vendor (TrioTim, Skyra models). For model testing, the Prostate 3T dataset [6] was used containing 30 patients and 421 frames in total with its corresponding annotations of the whole gland obtained in T2-Weighted MR images from Siemens vendor (Skyra model).

B. Deep learning models

The proposed meta-learner model was developed based on three well-established DL segmentation networks that have demonstrated state-of-the-art performance for segmenting the prostate.

U-net: The U-net [4] is a long-established architecture in the biomedical imaging domain for segmentation applications. The network consists of an encoder-decoder combination, where the encoder captures the semantic features regarding the area of interest while the decoder upsamples that information assisted by the skipped connections.

Dense2U-net: Dense2U-net [7] is an encoder-decoder network where, like residual connections, dense blocks [8] are being used to enable the passing of information from previous layers forward while the transitional blocks are reducing the computational burden of the network making the features produced by dense blocks simpler.

Bridged U-net: The Bridged U-Net [9] combines two interconnected U-net substructures, forming a W-shaped network, to transfer the information and further analyze the image into more complex patches. Feature maps are fused together using concatenation, thereby improving the performance of the network.

C. Proposed method

A meta-learner architecture was developed in order to select well refined features acquired from 3 well-established models dedicated for medical image segmentation tasks. Figure 1 shows the proposed meta-learner pipeline. The whole pipeline takes 2 inputs, which are the initial images along with their corresponding whole gland annotations, and each of them passes from each model. Few layers before each model's final decision, the feature maps are extracted, therefore only the most informative features for the decision of each network are kept. In addition, the concatenation process between the feature maps obtained by each individual model is ongoing, and the resulting feature map contains rich task-related information. This information is passed through the meta-learner module, which consists of convolutional filters with different kernel sizes seeking to harness close and distant information across the feature maps.

Specifically, Fig.1A shows the separation of the input in the models and the resulted feature maps from each network. The concatenation process of the resulted maps of Fig.1A is depicted in Fig.1B and described in Eq.(1):

$$FM_{dn+un+br} = \sum_{i=1}^{i=Nc_{dn}} FM_i + \sum_{i=1}^{i=Nc_{un}} FM_i + \sum_{i=1}^{i=Nc_{br}} FM_i \quad (1)$$

Where $FM_{dn+un+br}$ are the resultant concatenated feature maps, dn, un, br are the indices for Dense2U-net, U-net and Bridged U-net respectively, Nc are the number of channels extracted from each individual network and FM are the feature maps of each network.

In Fig.1C, the features extracted by Eq. (1) are passed through convolutional layers of different kernel size. This process enables the meta-learner to identify close spatially related features along with distant ones, and as a consequence, it is able to recognize patterns that individual models are not able to. $FM_{dn+un+br}$ maps are passed from the convolutions of different receptor field units and features from each operation are concatenated and then passed from a convolutional layer with a kernel of size 3x3. Finally they are passed from an identity convolutional layer and the sigmoid function to obtain the predicted prostate gland mask.

D. Evaluation metrics

For the evaluation of the models' performance 5 metrics have been used; Sensitivity in this particular study refers to a model's ability to identify correctly pixels belonging to the region of interest [10]. Balanced Accuracy (BA) is used when there is class imbalance and background pixels are prevalent to foreground [11]. Dice similarity Score (DS) measures the statistical similarity between 2 clusters of data points and is the most widely used and well-established overlap metric for medical image segmentation applications [12]. Hausdorff Distance (HD) is an one dimensional metric which measures how far two data points are (one from the ground truth and one from the predicted mask boundaries), by means of euclidean distance. Herein, the 95% HD was used, compensating for the impact of outliers [13]. Rand Error Rate (RER) measures the error of intensity correlation between two clusters of intensity values [14].

E. Model development

The backbone of the meta-learner consist the original U-net, Dense2U-net and Bridged U-net models while the model that has been added on top of the three base networks has 4 convolutions with different receptor fields. Batch normalization and specifically L2 regularization technique was utilized for gamma and beta regularizers with a value of 10^{-4}. The dropout rate was 0.5. Batch size was kept at 16, the number of epochs was set to 120, while early stopping was used to halt the training process when the model didn't improve further at the validation set. Binary cross-entropy was used as a loss function and Adam optimizer was employed for updating model's weights. The learning rate set to 10^{-4}. The GPU used for the experiments is an NVIDIA Quadro P6000 with drivers

Authorized licensed use limited to: Foundation for Research and Technology (FORTH). Downloaded on June 15,2023 at 08:40:01 UTC from IEEE Xplore. Restrictions apply.

TABLE I
MEAN VALUES AND STANDARD DEVIATION FOR 4 DEEP LEARNING MODELS ON THE TEST DATASET

	Sensitivity %	*Balanced accuracy %*	*Dice score %*	*Hausdorff distance (mm)*	*Rand error rate %*
Meta-learner	**0.78 ± 0.26**	**0.89 ± 0.13**	**0.79 ± 0.24**	20.94 ± 23.56	**0.17 ± 0.21**
Dense2U-net	0.66 ± 0.31	0.83 ± 0.15	0.71 ± 0.29	20.89 ± 21.64	0.21 ± 0.21
U-net	0.72 ± 0.31	0.86 ± 0.15	0.72 ± 0.28	23.42 ± 23.3	0.23 ± 0.27
Bridged U-net	0.59 ± 0.29	0.79 ± 0.14	0.67 ± 0.29	23.56 ± 23.39	0.27 ± 0.23

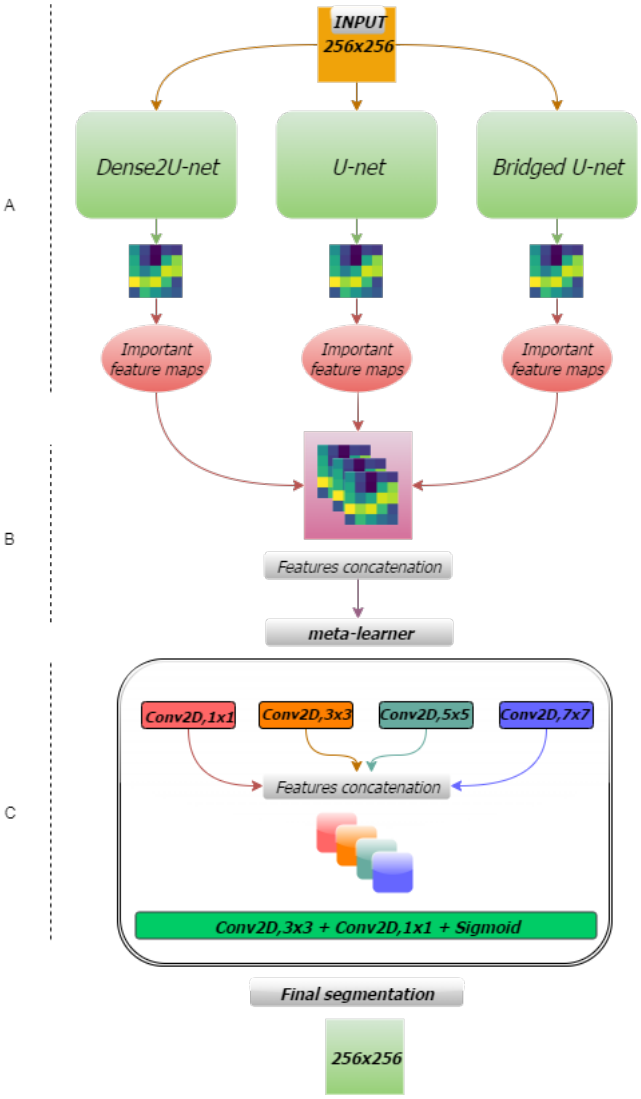

Fig. 1. The proposed meta-learner method

version 441.66. The python packages utilized were scikit-image=0.18.3, numpy=1.21.2, keras-unet-collection=0.1.11, scipy=1.7.1, tensorflow=2.2.0 and tensorflow-addons=0.11.2.

III. RESULTS

Table 1 presents the average evaluation scores obtained through cross-validation along with the corresponding standard deviation. The proposed meta-learner network achieved superior performance for all the evaluation metrics. With the exception of the HD metric, where the meta-learner and Dense2U-net performed equally well, the improvement achieved in sensitivity, BA, DS and RER compared to the best performing base-learners (either U-net or Dense2U-net) was 6%, 3%, 5%, and 4%, respectively.

Regarding the DS, the meta-learner achieved superior performance in terms of median value (89%) and interquartile range (IQR) [78%-93%] compared to the second best model, the U-net, which obtains median and interquartile values of 83% [IQR: 65%-90%]. Regarding the HD, the meta-learner shows greater performance in terms of median value (12 mm) and interquartile range (8-24 mm) compared to the second best model, the Dense2U-net, which obtains median and interquartile values of 14 mm and 9-28 mm.

Figure 2 depicts the overlaid ground truth and predicted contours by the models, in 6 different cases. Starting from the top cases 1,2,3 and 6 represent areas of the midgland prostate's region while case 4 is the apex and case 5 is the basal region. It is evident that for each individual model the segmentation of the prostate gland in apex and basal areas is much more challenging than the segmentation of the midgland region. Especially for the case 5, the individual models failed completely to recognize such a small cluster of pixels belong to the prostate, while the meta-learner performed relatively well. In fact, it takes a contradictory path and it is able to efficiently segment even the most challenging areas. Even in simpler cases (i.e Cases 1, 2, 3, and 6), the meta-learner was able to delineate more effectively the region of interest.

IV. DISCUSSION & CONCLUSION

In the present work, we propose a meta-learner network aiming to improve the segmentation accuracy of the prostate gland on pelvic MRI images. The innovative aspect of this work lies on the optimization of the feature extraction process using three individual DL models. The experiments suggest that the proposed model outperformed all the other networks in the majority of the metrics tested.

The meta-learner module practically employs the computational power of a well generalizable and simple model to identify features of low complexity (U-net) while it retains the ability of two related to the task models(Dense2U-net & Bridged U-net) to extract complex features. The features from each individual model are concatenated in channel-wise manner just before their passing from the convolution layer which is responsible for the extraction of the predicted mask. The concatenated features are then passed from the meta-learner model to extract valuable information regarding the area of interest from each model and cancel their individual errors.In the meta-learner, the feature weights are assigned

Authorized licensed use limited to: Foundation for Research and Technology (FORTH). Downloaded on June 15,2023 at 08:40:01 UTC from IEEE Xplore. Restrictions apply.

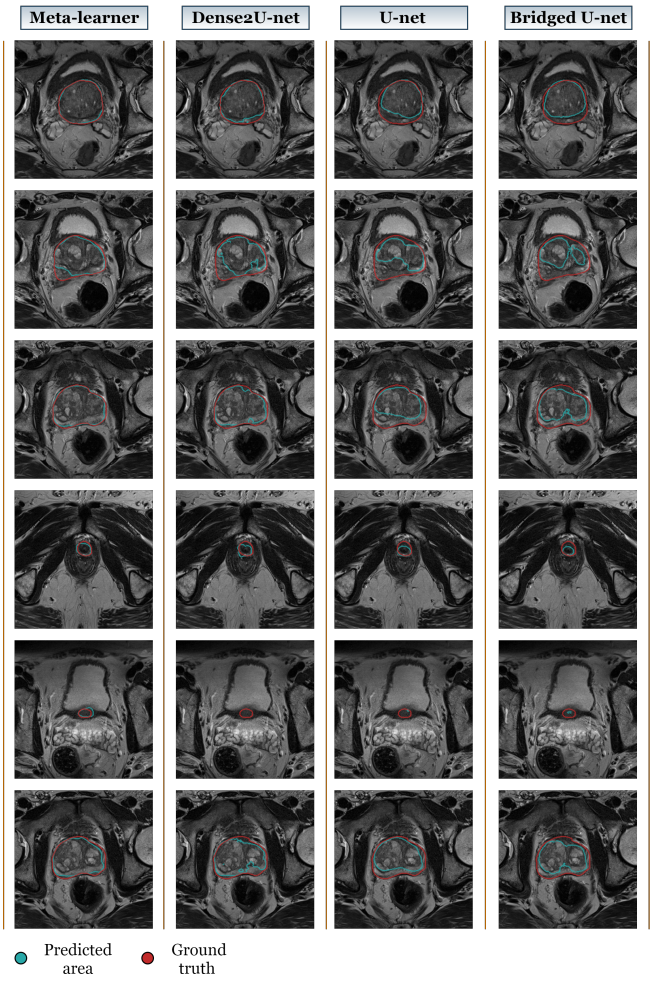

Fig. 2. The predicted (blue) and the ground truth (red) contours for 6 cases and 4 networks. Cases start top to bottom (1-6)

based on how close the base-learner's feature maps are to the ground truth. Thereby, base-learner's errors are minimized.

The main limitation of this study is that the experiment was conducted in relatively small amount of data. Future work could also take advantage of image preprocessing techniques to further improve the performance of the meta-learner. For instance, data augmentation is an efficient work-around when the amount of training data is limited. Also, using advanced techniques to crop the image around the prostate, in order to overcome issues related to the class imbalance [15], [16] between the prostate and the background regions, have been proven particularly useful for enhancing the performance of DL segmentation networks [17], [18].

ACKNOWLEDGMENT

This work is supported by the ProCancer-I project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 952159. It reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] S. E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R. E. Lenkinski, and A. Madabhushi, "Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo t2-weighted mr imagery," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 1, pp. 213–224, 2012.

[3] M. de Rooij, B. Israël, M. Tummers, H. U. Ahmed, T. Barrett, F. Giganti, B. Hamm, V. Løgager, A. Padhani, V. Panebianco, P. Puech, J. Richenberg, O. Rouvière, G. Salomon, I. Schoots, J. Veltman, G. Villeirs, J. Walz, and J. O. Barentsz, "Esur/esui consensus statements on multiparametric mri for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training," *European Radiology*, vol. 30, pp. 5404–5416, Oct 2020.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv*, 2015.

[5] R. Cuocolo, A. Stanzione, A. Castaldo, D. R. De Lucia, and M. Imbriaco, "Quality control and whole-gland, zonal and lesion annotations for the prostatex challenge public dataset," *European Journal of Radiology*, vol. 138, no. 109647, 2021.

[6] G. Litjens, J. Futterer, and H. Huisman, "Data from prostate-3t. the cancer imaging archive.," 2020.

[7] N. Aldoj, F. Biavati, F. Michallek, S. Stober, and M. Dewey, "Automatic prostate and prostate zones segmentation of magnetic resonance images using densenet-like u-net," *Scientific Reports*, vol. 10, no. 1, pp. 2045–2322, 2020.

[8] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Los Alamitos, CA, USA), pp. 2261–2269, IEEE Computer Society, jul 2017.

[9] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, and X. Tang, "W-net: Bridged u-net for 2d medical image segmentation," *arXiv*, 2018.

[10] R. Trevethan, "Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice," *Frontiers in Public Health*, vol. 5, 2017.

[11] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *2010 20th International Conference on Pattern Recognition*, pp. 3121–3124, 2010.

[12] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports," *Academic Radiology*, vol. 11, no. 2, pp. 178–189, 2004.

[13] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[14] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pp. 34–34, 2005.

[15] E. Tappeiner, M. Welk, and R. Schubert, "Tackling the class imbalance problem of deep learning based head and neck organ segmentation," *CoRR*, vol. abs/2201.01636, 2022.

[16] D. G. Zaridis, E. Mylona, N. S. Tachos, K. Marias, N. Papanikolaou, M. Tsiknakis, and D. I. Fotiadis, "A new smart-cropping pipeline for prostate segmentation using deep learning networks," 2021.

[17] D. Zaridis, E. Mylona, N. Tachos, K. Marias, M. Tsiknakis, and D. I. Fotiadis, "A deep learning-based cropping technique to improve segmentation of prostate's peripheral zone," in *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 1–4, 2021.

[18] D. Zaridis, E. Mylona, N. Tachos, K. Marias, M. Tsiknakis, and D. I. Fotiadis, "A smart cropping pipeline to improve prostate's peripheral zone segmentation on mri using deep learning," *EAI Endorsed Transactions on Bioengineering and Bioinformatics*, vol. 1, 2 2022.

Authorized licensed use limited to: Foundation for Research and Technology (FORTH). Downloaded on June 15,2023 at 08:40:01 UTC from IEEE Xplore. Restrictions apply.