3D Multi-eXpert Fusion Framework For Prostate Segmentation In 3D TRUS Imaging

Clément Beitone, Jocelyne Troccaz

3D Multi-eXpert Fusion Framework For Prostate Segmentation In 3D TRUS Imaging

Clément Beitone
Univ. Grenoble Alpes, CNRS, UMR 5525
VetAgro Sup, Grenoble INP, TIMC
38000 Grenoble, France
0000-0001-5921-9419

Jocelyne Troccaz
Univ. Grenoble Alpes, CNRS, UMR 5525
VetAgro Sup, Grenoble INP, TIMC
38000 Grenoble, France
0000-0003-4081-4247

***Abstract*— Prostate segmentation is a key step for several clinical applications, both preoperatively and intraoperatively. Efforts to automate it would contribute to a more consistent level of quality for this procedure and reduce inter- and intra-operator variability, thus ensuring better patient care.**

We introduce 3D-MXF an automatic segmentation approach using ensemble deep learning to merge the segmentation result proposed by several expert networks. Three CNN are trained to specifically segment 2D images extracted from a single 3D TRUS volume according to the three main views (axial, coronal, sagittal). The main contribution lies in the specific fusion step performed by a 3D CNN trained to provide 3D confidence maps that enable the fusion of the volumes reconstructed from the expert networks' outputs into a final segmented volume.

The 3D-MXF framework was trained and carefully evaluated on a database containing prostate TRUS images of 382 patients. The results are superior to what is obtained with other state-of-the-art methods. A fine analysis of clinical parameters demonstrated the potential and limitations of this approach for a clinical use.

***Index Terms*—Prostate segmentation, 3D TRUS, Deep Learning, Ensemble learning.**

I. Introduction

Despite continuous advances in therapy and systematic screening campaigns, prostate cancer remains one of the leading causes of death in men [13]. The reference method for screening and therapeutic follow-up is the trans-rectal prostate biopsy performed under ultrasound guidance. On the latest computer-assisted ultrasound platforms, the biopsy process starts with the segmentation of the prostate. The accuracy of this segmentation is crucial as it directly affects the quality of guidance provided by the platform software during the biopsy procedure [8]. Implementing a reliable, precise and fast method would enhance patient care. Additionnally, reducing the required time could minimize patient discomfort during the biopsy session. Lastly, a reliable automatic approach could help reduce the inter- and intra-operator variability that is now an established concern [14].

Numerous techniques for prostate segmentation have emerged in recent years [12]. However, the inherent complexity of this task presents a formidable challenge, even for state-of-the-art deep learning algorithms operating on extensive datasets. One of the primary challenges stems from the distinct anatomical characteristics unique to each prostate, which lead to varying levels of echogenicity in images across different patients. Furthermore, it is widely recognized that delineating the basal and apical regions, due to their positioning and proximity to other tissues and organs, poses a significant challenge, even for experienced urologists

Many of the methods presented in the existing literature predominantly depend on 2D techniques, which often exhibit reduced accuracy and reliability in the basal and apical regions. This deficiency arises from the absence of consistency that a third dimension would offer. On the other hand, 3D methodologies confront the challenge of managing extensive data volumes necessary to train more complex convolutional networks. This factor also influences the need for datasets that must be not only sufficiently large but also representative of the diverse patient population [4].

An intriguing solution that combines the advantages of leveraging a vast amount of 2D data while ensuring the segmentation accounts for three-dimensional spatial coherence would be of significant interest. Similar strategies have been explored in the domain of machine learning, particularly within the context of ensemble learning [6]. In this approach, multiple so-called *experts*, each specializing in specific aspects of the dataset, are trained. None of these experts, alone, is able to solve the problem posed by the whole database. An algorithm is then necessary to combine the answers of these different experts and propose a consensual answer to the overarching problem. Other approaches such as STAPLE or linear fusion [11] have been proposed but their superiority as fusion algorithms has not been demonstrated in all cases. Recent research [7, 1] has put forth the concept of using a specially trained network to optimally manage this fusion task. The fundamental idea here is that each expert can offer valuable insights into specific regions of the image or volume, and training the network allows for assigning variable and local confidence levels to these diverse experts. Regrettably, these approaches share a common drawback. They treat volumes as slices and reconstruct the 3D segmentation by merely stacking the segmented slices. Consequently, the algorithm cannot effectively capture the spatial coherence present along the reconstruction axis.

The primary contribution of this study is the introduction

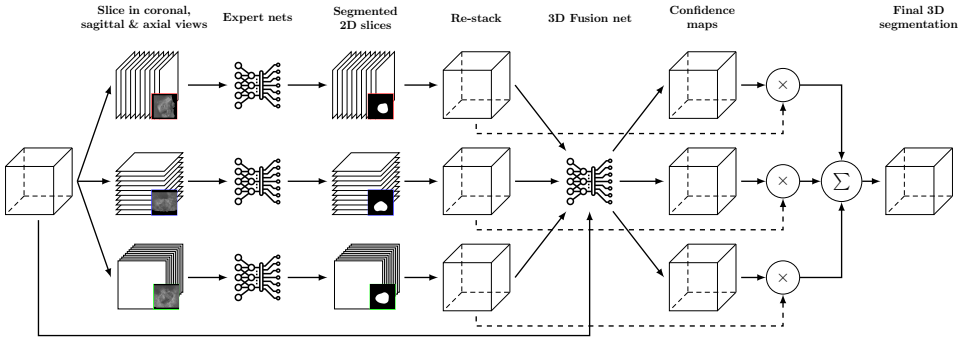

Figure 1. 3D-MXF framework overview.

of a novel framework that addresses the limitations identified in the prior work 2D-MXF [1]. Specifically, this framework entails a three-dimensional fusion of the segmentation maps generated by the expert networks. Through the integration of a new 3D fusion network, this approach accounts for the spatial coherence among the slices segmented by the expert networks. The secondary contribution of this research lies in the assessment of the proposed 3D-MXF. A meticulous examination has been conducted to assess the significance of the obtained results. This evaluation extends beyond the analysis of conventional image analysis metrics and also incorporates factors such as the reconstructed volume, which aligns with a more clinical perspective.

II. Method

A. 3D-MXF Framework

3D-MXF (see Figure 1) is designed to provide an estimate $\hat{Y}$ of the reference 3D segmentation Y that an urologist produces from 3D TRUS imaging of a prostate. To achieve this outcome, and given that the proposed approach can be framed within the context of ensemble learning, two steps are required. Firstly, three distinct convolutional networks, refered to as *expert networks* (see Section II-B) are trained to segment 2D ultrasound slices extracted in the three main views of the original 3D volume (sagittal, coronal and axial).

Segmentating all the 2D slices within each of the three orientations using the expert networks allows to obtain, after re-stacking, three segmented volumes noted E_i (with $i \in \{sagittal, coronal, axial\}$). These volumes alongside the original 3D TRUS volume are subsequently fed to a *fusion network* (cf. Section II-C) which, in turn, produces three 3D confidence maps denoted as C_i. When combined with their respective segmented volume E_i,these confidence maps form the segmented volume $\hat{Y}$. The fusion process, by means of these maps, assigns a local confidence, at the voxel level, to each of the results generated by the expert networks. This results in a final segmentation that is more robust, as it results from this mixture of information from several sources. Equation 1 describes the operations necessary for constructing the $\hat{Y}$ estimate as well as the constraint that must be imposed on the confidence maps.

$$\hat{Y} = \sum_{i=1}^{3} C_i(X) E_i(X), \text{ with } \begin{cases} 0 \leq C_i(X) \leq 1 \\ \sum_{i=1}^{n} C_i(X) = 1 \end{cases} \quad (1)$$

B. Expert networks

The experts of the 3D-MXF approach are 2D U-Nets. A U-Nets consists in an encoding branch followed by a decoding branch with skip-connections. The encoding and decoding blocks are series of two 3×3 convolutional layers, each followed by batch normalization and ReLU. Each encoding block doubles the number of features and is connected by 2×2 max-pooling layers. The decoding blocks halve the number of features and are connected by 2×2 max-pooling layers. The output is generated by a 1×1 convolution layer followed by a sigmoid unit. The network and loss have been trained with the Adam optimizer with a fixed learning rate set at $1e^{-4}$ to optimize a Dice loss function. Orlando *et al.* [10] have shown the efficiency of the Dice loss for prostate TRUS image segmentation.

$$Loss_{Dice}\left(Y, \hat{Y}\right) = 1 - \left(2 \cdot \frac{|Y \cap \hat{Y}|}{|Y| + |\hat{Y}|}\right) \quad (2)$$

C. Fusion network

To perform the fusion, the 3D-MXF framework relies on a 3D U-Net. A 3D U-Net is a 2D U-Net (see Section II-B) where 2D convolution layers are replaced by 3D convolution layers. The input of this fusion network is a 4D tensor comprising the three volumes produced by the expert networks and the original ultrasound volume. The network is trained to compute three 3D confidence maps with values between 0 and 1 (a range ensured by the use of a softmax layer at the network output).

The fusion network is trained using a combination of a Dice loss and a loss based on an estimate of the length of the segmented contour. This length loss function used in [2] allows to penalize non-smooth contours (see Equation 3):

$$Loss_{Length}\left(\hat{Y}\right) = \sum_{\Omega} \sqrt{\left|\left(\nabla \hat{Y}_{x_{i,j,k}}\right) + \left(\nabla \hat{Y}_{y_{i,j,k}}\right) + \left(\nabla \hat{Y}_{z_{i,j,k}}\right)\right| + \epsilon} \quad (3)$$

Table I
COMPARISON WITH THE 2D-MXF, THE STAPLE STRATEGIES AND A V-NET ARCHITECTURE.

Crit.	3D-MXF	2D-MXF	STAPLE	VNET
Dice	$0.93 \pm .02$	$0.92 \pm .02$	$0.89 \pm .02$	$0.90 \pm .02$
Jaccard	$0.87 \pm .04$	$0.86 \pm .04$	$0.84 \pm .04$	$0.85 \pm .04$
ASD	$0.84 \pm .31$	$0.96 \pm .34$	$1.21 \pm .42$	$1.11 \pm .39$
HD	4.32 ± 1.52	5.35 ± 2.22	8.75 ± 5.57	7.15 ± 4.57

where $\nabla_{\{x/y/z\}_{i,j,k}}$ corresponds to the derivatives in each of the three dimension.

D. Dataset and ground truth

The database used to train and evaluate the proposed approach consists of 3D TRUS images acquired during biopsy sessions at the University Hospital of Grenoble, conducted for the suspicion or follow-up of prostate cancer over a five-year period. The TRUS volumes were acquired with two different machines (Trinity® platform and Urostation®) including motorized 3D end-fire TRUS probes.

The images underwent segmentation by various expert urologists during the biopsy examinations. Ground truth data were generated through a semi-automatic segmentation process carried out with the ultrasound platform. Prior to their inclusion in the training database, each data element was additionally reviewed by a human expert to guarantee the quality of the semi-automated segmentations. Patients whose segmentations exhibited noticeable quality defects were excluded from the database. Following this thorough evaluation, a total of 382 patients remained for training and validating the proposed approach.

E. Evaluation

The results obtained through the proposed approach were compared with several state-of-the-art methods. First, a direct segmentation approach based on a 3D V-Net model [9] was implemented to facilitate a comprehensive evaluation. Furthermore, two fusion strategies, majority voting and the staple [16] method, were implemented to fuse the results produced by the three expert networks, as well as the 2D-MXF fusion approach. All these methods were evaluated using a k-fold cross-validation method on the database. We chose k=5 to have a ratio of 80% of volumes for training and 20% for testing.

F. Implementation

The proposed framework and the reference networks presented in Section II-E were implemented using the Pytorch framework. The experiments were conducted on a server equipped with NVIDIA A100 GPUs. All networks underwent separate training : the three expert networks were trained first, followed by the fusion network. Each network was trained for 100 epochs. The batches consisted of 50 2D images for the expert networks. The fusion network used batches of 3 volumes.

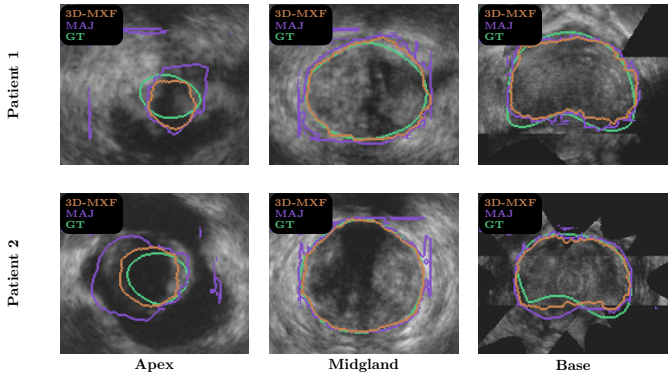

Figure 2. Contours obtained with the 3D-MXF method (orange) and the MAJ fusion method (purple) for two different patients (top and bottom rows) and at various prostate levels (basal, midgland, and apical regions).

III. RESULTS

The values shown in Table I for the 3D-MXF, 2D-MXF, STAPLE and V-Net strategies correspond to the results obtained through the k-fold validation methodology. First, we observe, in agreement with Hatt *et al.* [5], that STAPLE yields results identical to those of MAJ (and are therefore not included in the Table I). Furthermore, this table clearly demonstrates that the MXF strategies, whether in 2D or 3D, notably enhance the segmentation quality when compared to the other conventional approaches outlined here. It can also be seen that taking into account the third dimension in the 3D-MXF framework improves the result compared to the original 2D-MXF approach. This difference can be explained by the fact that the re-stacking reconstruction in the 2D-MXF method does not guarantee the continuity along the reconstruction axis. The performance differences between the 3D-MXF method and the other approaches are all statistically significant (Shapiro / Wilcoxon tests). In terms of metrics, the advantages of the proposed approach become evident when considering that, while the overlap indices (Dice and Jaccard) are relatively close, indicating the capability of different methods to deliver a similar overall result, the principal contribution of the proposed method lies in its ability to yield a relatively smooth contour with very few outliers, resulting in lower Hausdorff and ASD values.

When comparing the results of the expert networks alone to the proposed approach, the Dice and Jaccard indices are improved by 10% on average. This improvement is even more visible when analyzing the ASD values and the Haussdorff index which are respectively decreased by 30% and 55% on average.

Figure 2 shows more precisely on different slices extracted from two patients typical differences between contours obtained with the 3D-MXF method and the MAJ method. This figure also shows that the most important difference appears in the extreme regions (apex & base).

The results of the 3D-MXF approach can also be compared to the state of the art. Wang *et al.* [15] obtained an average Dice of 0.90 ± 0.02 (Jaccard 0.82 ± 0.04) and an ASD of

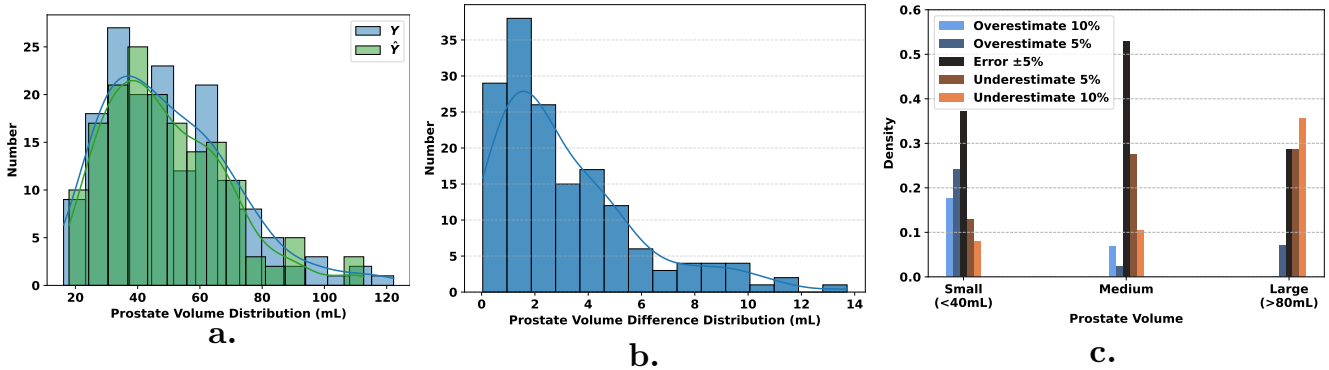

Figure 3. Prostate Volume Analysis. (a) Histograms of volumes derived from manual and 3D-MXF automated segmentation. (b) Histogram of absolute volume differences between ground truth and predictions. (c) Density distribution of overestimation and underestimation in prostate volume using the 3D-MXF approach for small, medium, and large prostates.

1.16 ± 0.7 mm in the best cases on a dataset composed of 40 patients (with a four-fold cross-validation strategy, i.e. 10 test patients for each fold). Orlando *et al.* [10] obtained a median Dice of 0.94 and a median ASD of 0.89 mm on their test dataset composed of 20 patients. Girum *et al.* [3] proposed a 3D segmentation method based on 2D segmentation but without any fusion step. They reported a mean Dice of 0.88 ± 0.02 and a mean HD of 8.37 ± 2.93 mm across 14 patients. It is worth noting that the evaluation of these methods was performed on considerably smaller patient datasets compared to our own database.

Figure 3a. depicts the distribution of volumes obtained on the test dataset. In our database, the three categories of prostate size (small, medium, large) account for 38, 53 and 9% of the cases, respectively. On the test database, the average volume of the prostate as manually segmented by the clinicians is 50 mL ($\pm$ 21) while it measures 51 mL ($\pm$ 22) with the 3D-MXF approach. The average difference is 3.7 mL ($\pm$ 4.3) and most of the errors fall within a range of less than 6 mL. Figure 3b. presents the distribution in prostate volume differences obtained across the entire database.

Figure 3c. illustrates the ratio of overestimation and underestimation for the three categories of prostate size. We consider that there is no underestimation or overestimation if the estimated volume is within 5% of the ground truth volume [8]. The figure reveals that the method tends to more frequently underestimate the prostate volume for cases with large prostates. This observation may be attributed to two factors: firstly, large prostates constitute only 9% of the cases in the database. As a result, our model is less exposed to these cases and tends to produce results closer to the average of the most prevalent class, i.e., medium-sized prostate. Moreover, given the lower representation of large prostates in the population, clinicians' ability to accurately segment these larger prostates may also vary. This variability in clinician performance contributes to tempering the errors made by the network, which tends to underestimate prostate size towards the mean value as inter-operator variability increases.

IV. Conclusion

This work focuses on the segmentation of 3D TRUS images of the prostate with the aim of replacing the manual or semi-automatic segmentation steps that are currently a part of many clinical procedures. These steps are susceptible to significant inter- and intra-operator variability and are time-consuming

The 3D-MXF framework we propose combines the outputs of the three expert networks to create a consensus segmentation result that is more robust compared to what each individual expert network can provide. One of the key contributions of this work is that the fusion network is not trained to generate a segmentation estimate, but rather to produce confidence maps. These confidence maps are used to reconstruct the final result by integrating them with the segmentations estimated by the expert networks

Through the evaluation results, this approach has demonstrated its capacity to generate smooth contours that closely align with clinicians' expectations. This is particularly evident in the apical and basal regions, where methods like majority voting tend to yield subpar results.

From a qualitative perspective, our approach stands out favorably in comparison to all the methods presented in the current state of the art. It is crucial to note that our approach's evaluation is conducted on a database that is more than three times larger in size than the most advanced methods found in the literature.

Our database includes all patients examined over a 5-year period, resulting in significant diversity, exemplified by the substantial variation in prostatic volumes, ranging from 1 to 6. Within this database, one can observe the efficacy of the proposed approach in delivering solutions with an average deviation of just 3.7 mL to the ground truth values. Evaluating the model in terms of prostate volume categories underlines the significance of training deep learning methods on databases that closely resemble those encountered in everyday clinical practice. The algorithm makes systematic errors in cases that are under-represented but these errors remain limited. One possible approach would be to target each of these sub-groups and apply a tailored model to them.

V. ACKNOWLEDGMENTS

This work was partly supported by Auvergne Rhône-Alpes region (project ProNavIA) and by the French "Agence Nationale de la Recherche", Investissement d'Avenir program (grants MIAI@Grenoble-Alpes under reference ANR-19-P3IA-0003 and CAMI Labex under reference ANR-11-LABX-0004).

REFERENCES

[1] Clément Beitone and Jocelyne Troccaz. "Multi-eXpert fusion: An ensemble learning framework to segment 3D TRUS prostate images". In: *Medical Physics* 49.8 (2022), pp. 5138–5148.

[2] Xu Chen et al. "Learning active contour models for medical image segmentation". In: vol. 2019-June. IEEE Computer Society, June 2019, pp. 11624–11632. ISBN: 9781728132938.

[3] Kibrom Berihu Girum et al. "A deep learning method for real-time intraoperative US image segmentation in prostate brachytherapy". In: *International Journal of Computer Assisted Radiology and Surgery* (2020). ISSN: 18616429.

[4] Grant Haskins, Uwe Kruger, and Pingkun Yan. "Deep learning in medical image registration: a survey". In: *Machine Vision and Applications* (2020). ISSN: 14321769.

[5] Mathieu Hatt et al. "The first MICCAI challenge on PET tumor segmentation". In: *Medical Image Analysis* 44 (Feb. 2018), pp. 177–195. ISSN: 13618423.

[6] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2 (2021), pp. 203–211. ISSN: 15487105.

[7] Mingjie Jiang, J David Spence, and Bernard Chiu. "Segmentation of 3D ultrasound carotid vessel wall using U-Net and segmentation average network". In: 2020, pp. 2043–2046.

[8] Louis Lenfant et al. "Impact of Relative Volume Difference Between Magnetic Resonance Imaging and Three-dimensional Transrectal Ultrasound Segmentation on Clinically Significant Prostate Cancer Detection in Fusion Magnetic Resonance Imaging-targeted Biopsy". In: *European Urology Oncology* (Aug. 19, 2023). ISSN: 2588-9311. DOI: 10.1016/j.euo.2023.07.016. URL: https://www.sciencedirect.com/science/article/pii/S2588931123001608 (visited on 10/04/2023).

[9] Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. "V-Net: Fully convolutional neural networks for volumetric medical image segmentation". In: *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016* (2016), pp. 565–571.

[10] Nathan Orlando et al. "Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images". In: *Medical Physics* 47.6 (2020), pp. 2413–2426. ISSN: 00942405.

[11] Mathias Perslev et al. "One Network to Segment Them All: A General, Lightweight System for Accurate 3D Medical Image Segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11765 LNCS (2019), pp. 30–38. ISSN: 16113349.

[12] J Ramesh et al. *Review on Computational Models for Prostate Segmentation from Ultrasound Medical Images Review Article*. 2022.

[13] Hyuna Sung et al. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA: A Cancer Journal for Clinicians* 71.3 (2021), pp. 209–249. ISSN: 0007-9235.

[14] Shidong Tong et al. "Intra-and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging". In: *Ultrasound in medicine & biology* 24.5 (1998), pp. 673–681.

[15] Yi Wang et al. "Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound". In: *IEEE Transactions on Medical Imaging* 38.12 (2019), pp. 2768–2778. ISSN: 1558254X.

[16] Simon K Warfield, Kelly H Zou, and William M Wells. "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation". In: *IEEE transactions on medical imaging* 23.7 (2004), pp. 903–921.