EFFICIENT CATHETER SEGMENTATION IN 3D CARDIAC ULTRASOUND USING SLICE-BASED FCN WITH DEEP SUPERVISION AND F-SCORE LOSS

Hongxu Yang**, Caifeng Shan*†*, Alexander F. Kolen*†*, Peter H.N. de With**

*Eindhoven University of Technology, Eindhoven, The Netherlands
†Philips Research, Eindhoven, The Netherlands

ABSTRACT

Fast and accurate catheter segmentation in 3D ultrasound (US) can improve the outcome and efficiency of cardiac interventions. In this paper, we propose an efficient catheter segmentation method based on a fully convolutional neural network (FCN). The FCN is based on a pre-trained VGG-16 model, which processes the 3D US volumes slice by slice. To enhance its performance, we modify its structure by skipping connections under a deep supervision structure, which is learned with an F-score loss function. Our method can exploit more contextual information and increase the detection of catheter-like voxels. We collected a challenging *ex-vivo* dataset (92 3D US images) from porcine hearts with an RF-ablation catheter inside. Our experiments on this dataset show that the proposed method achieves a segmentation performance with an F_2 score of 65.2% with a highly efficient inference around 1.1 sec. per volume.

Index Terms— 3D ultrasound, catheter segmentation, FCN, F-score loss, deep supervision.

1. INTRODUCTION

During cardiac interventions being applied at large scale, medical imaging methods like fluoroscopy and Ultrasound (US) imaging have been employed to visualize obscured instruments or tissue. Because of the maturity of 3D cardiac US imaging, 3D US has recently become an attractive option to offer richer spatial information than conventional X-ray to guide the surgery. More crucially, 3D US is radiation-free for both patient and surgeon. However, due to the low resolution and low contrast of 3D US images, it is difficult for the sonographer to localize the catheter in the 3D US. Therefore, automatic catheter detection or segmentation of instruments in general is highly desired for enhancing clinical applications [1].

Various techniques for image-based catheter detection have been explored. Conventional methods like template matching, projection-based line fitting, or handcrafted features with voxel-based supervised machine learning methods have been proposed and have shown promising results using different datasets [2][3][4][5][6][7]. However, the above methods have still several drawbacks and challenges. Template matching appears to be not stable when the image appearance has large variations or too much noise [3]. As for handcrafted-feature methods, the limited representation capability cannot always handle complex 3D US images containing anatomical structures [4][5][6][7]. More recently, deep learning methods, such as convolutional neural networks (CNNs), have achieved a large performance improvement in the medical imaging [8]. For medical instrument detection or segmentation in 3D US, two approaches have been exploited: voxel-based classification [9][10] and slice-based semantic segmentation [11][12][13]. Although performances have been improved, there are still limitations in the existing deep learning approaches. For voxel-wise processing, the CNN is used to label the 3D image in a voxel-by-voxel way, which cannot fully exploit context information and may lead to high computation cost [9][10]. For slice-based semantic segmentation, methods employ 2D fully convolutional networks (FCN) to annotate the instruments' labels simultaneously on decomposed 2D slices from 3D US volumes. Although this approach provides efficient inference for 3D US images, the proposed structure [11][12][13] cannot fully exploit semantic information, as the network becomes deeper in its layering [14].

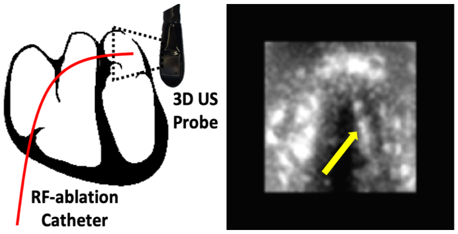

Fig. 1. Left: Example of catheter capturing using a 3D cardiac US probe during intervention. Right: Example of US image with catheter visibility inside (pointed by yellow arrow).

To address the problems in efficiency and catheter segmentation in 3D US images, we propose a novel segmentation method with the following aspects. (1) For efficiency, we consider a slice-based FCN, which is modified with skip connections to fuse information at different image scales. (2) We introduce a pyramidal deep supervision scheme, guiding the

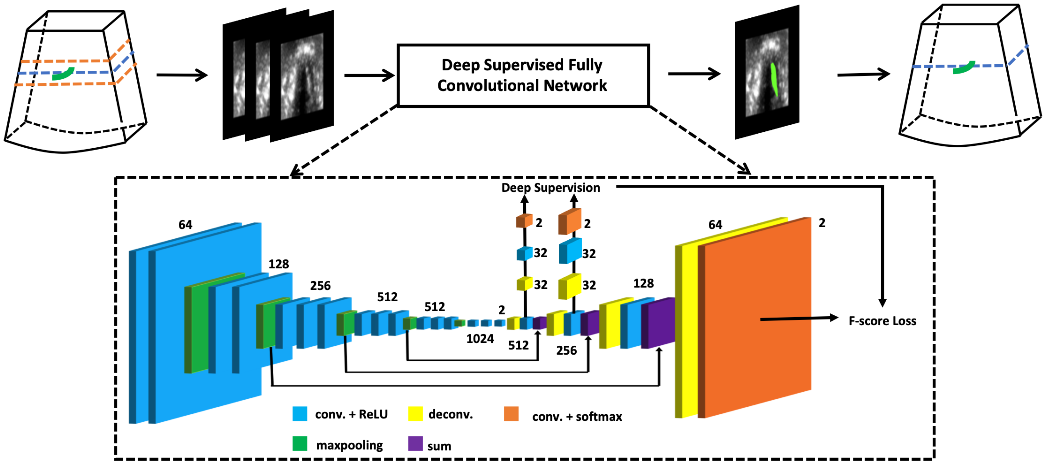

Fig. 2. Proposed catheter segmentation method. The input 3D volume is decomposed into adjacent 2D slices along the axial axis, as shown at the top. The target slice with its adjacent slices are supplied to the 2D transfer-learned FCN to obtain segmentation, which is trained by using a deep supervised F-score loss with a pre-trained VGG-16 model.

network to learn meaningful information at different scales. (3) We propose an F-score type loss function, which concentrates the network on successfully finding and segmenting the catheter from challenging US images, while improving the true positive rate (TPR). For validation, we have collected challenging *ex-vivo* 3D cardiac US datasets, on which our method achieves an F_2 score at 65.2% with a highly efficient inference around 1.1 second per volume.

2. OUR METHOD

Fig. 2 depicts the structure of the proposed catheter segmentation using slice-based FCN in 3D US. The 3D US volume is decomposed into slices along the axial axis. For each slice at the center, two adjacent slices with distance d are extracted ($d = 2$ in this paper). The three images are formed to be a 'fake' RGB image, as shown in Fig. 2 (top), where the center blue image serves as reference. The fake RGB image is then fed into a 2D FCN, which is shown in Fig. 2 (bottom). The FCN output, corresponding to the reference center image, is stacked back to its original position in the 3D volume. The final prediction on the 3D volume is obtained by stacking all the slice-based predictions along the axial direction. The reason for choosing the axial direction is twofold. First, the US probe is being placed by nature parallel to the heart chamber, and second, the axial direction has the worst spatial resolution compared to the remaining two directions. More details of the semantic segmentation are given below.

2.1. Slice-Based FCN with Deep Supervision

Our designed FCN is based on the VGG-16 structure [15]. As is shown in Fig. 2, VGG-16 has convolutional layers with 5-level max-pooling operations. After the last max-pooling layer, 3 convolutional layers are followed with kernel numbers 1024, 1024 and 2. Then, 4 deconvolutional layers are followed, which have filter size of 2×2, 2×2, 2×2 and 4×4, respectively. Moreover, for each deconvolutional layer, an additional convolution operation is added to improve the stability. To fuse the discriminating information at different scales, we add skipping connections like U-Net [12], which are shown at the bottom in Fig. 2. To further improve the performance of FCN, deep supervision is employed at different feature scales at the decoder: as shown in Fig. 2, the deconvolutional operation is first applied to upscale to input size, two convolutional layers are followed to concentrate information. For both deep supervision and the normal output layer (orange block), the final output is the probability mapping from the softmax operation, which is used as prediction or objective loss function. This function is discussed later.

In the training, the parameters of the encoding part in FCN are initialized based on the pre-trained VGG-16 model [15]. The remaining parameters are initialized randomly. To generalize the network, data augmentation techniques like rotation, mirroring, contrast transformation and scaling are applied. To augment the training images, each annotated catheter voxel in ground truth is used as the center of the 'fake' RGB image, which introduces a translation invariance in a natural way to facilitate catheter segmentation. Moreover, to enforce the network to learn the anatomical structure of the heart, non-catheter slices, i.e., slices consider non-catheter voxel as the center point, are downsampled to the same size as the catheter voxels to generate some images without catheter inside. All parameters of the FCN are learned by minimizing the F-score loss, using an Adam optimizer with a learning rate of 10^{-5}. Furthermore, to avoid overfitting, dropouts with probability 0.85 are introduced after convolution in the bottleneck layers (Layers 14 and 15 in VGG).

2.2. Loss Function

The commonly used loss function is the Dice loss [16] in medical image tasks. The Dice coefficient is specified by

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \, , \tag{1}$$

where TP denotes true positive, FN is false negative, and FP stands for false positive. However, when segmented images have an imbalanced class distribution and low and/or noisy image quality, the segmentation performances may not be optimal, in the sense that the prediction result tends to suppress the catheter voxels leading to a lower true positive rate. As discussed in our previous work [10], a higher true positive rate will fuel a successful catheter segmentation, while false positive voxels can be removed by post-processing [1].

To address the above considerations and achieve a higher true positive performance, we employ the F-score like loss function. The F-score is given by

$$\text{F-score}(\beta) = \frac{(1+\beta^2)TP}{(1+\beta^2)TP + \beta^2 FN + FP} \, , \tag{2}$$

where parameter β is used to control the learning weight between catheter voxels and non-catheter voxels. In this definition of F-score, when $\beta > 1$, the loss function will emphasize the true positive rate which would improve the Recall performance, while when $\beta < 1$, the network increases the Precision performance. When $\beta = 1$, the F-score equals the Dice function as given in Eqn. (1).

A further step to control the learning is to derive an indication for a softmax-based probability. To this end, we replace the parameters TP, FN and FP by voxel-based parameters and related probabilities for both catheter voxels and non-catheter voxels. Following this and more specifically, Eqn. (2) can then be rewritten into

$$F\text{-score}(\beta) = \frac{(1+\beta^2) \cdot \sum_{i=1}^{N} y_{ci}\hat{y}_{ci}}{(1+\beta^2) \cdot \sum_{i=1}^{N} y_{ci}\hat{y}_{ci} + \beta^2 \sum_{i=1}^{N} y_{ci}\hat{y}_{ni} + \sum_{i=1}^{N} y_{ni}\hat{y}_{ci}} . \tag{3}$$

In Eqn. (3), parameter y_{ci} denotes a catheter voxel from the ground truth, $\hat{y}_{ci}$ represents the voxel's probability of the prediction for that catheter class, while y_{ni} and $\hat{y}_{ni}$ are a non-catheter voxel and its corresponding probability, respectively. Based on F-score(β), the network can be controlled to learn the parameters while focusing on obtaining a higher Recall or Precision.

Based on the previous definitions and deep supervision mentioned earlier, our objective loss function $L(.)$ is now defined by

$$\begin{aligned} L(\beta, y, \hat{y_0}, \hat{y_1}, \hat{y_2}) =& F\text{-score}(\beta, y, \hat{y_0}) + \alpha_1 F\text{-score}(\beta, y, \hat{y_1}) \\ &+ \alpha_2 F\text{-score}(\beta, y, \hat{y_2}), \end{aligned} \tag{4}$$

where y is the ground truth, $\hat{y_0}, \hat{y_1}, \hat{y_2}$ are predictions from convolutional feature mappings after deconvolutional layers 64, 256 and 512 respectively, which are shown in Fig. 2 orange blocks. Constants α_1 and α_2 are loss weights, which are empirically determined as 0.5 and 0.25, respectively..

2.3. Post-processing

With slice-based FCN along the axial axis, the predicted slices are stitched back to their original position to obtain the final prediction.The cylinder-shape catheter in the segmented US can be detected by model-fitting algorithms from noisy segmentation, which are thoroughly studied in [1]. In this paper, we only consider the simplest post-processing method, i.e., consider the largest connective group in the segmented volume as the catheter. It requests a high performance on segmentation, requiring the Recall to be as high as possible.

3. DATASETS AND EXPERIMENTS

We have collected 92 3D cardiac US images from isolated porcine hearts. During the recording, the hearts were placed in water tanks with an RF-ablation catheter inside the chamber (diameter ranging within 2.3-3.3 mm). Because of the variation in recording conditions and setups, the volume size is ranging from $120 \times 69 \times 92$ to $294 \times 283 \times 202$ voxels, in which the voxel size was isotropically resampled to the range of 0.4-0.7 mm. Ground truth was obtained by voxel-based indications, manually annotated by clinical experts, which form binary masks indicating the voxels belonging to the catheter.

In our experiments, the dataset was randomly divided into 62 volumes for training and 30 volumes for testing. The voxel size for all voxels is not resampled into the same scale, to avoid possible artifacts from too much interpolation or downsampling. In the training, for each training volume, training images were extracted based on the sampling strategy mentioned in Section 2.1, which were slices along the axial axis. In the testing, the test volume was decomposed into slices and was predicted slice by slice. For simplicity, we selected $d = 2$ for all experiments. To evaluate the segmentation performance, Recall, Precision, and F_2 score per image are considered as metrics. We firstly perform an ablation study to show the effects of the deep supervisions scheme and the F-score loss. Then we compare it with the state-of-the-art catheter or medical instrument segmentation methods on our challenging datasets.

3.1. Ablation Study

This section describes the results of an ablation study, in which several cases of different structures of FCN are analyzed. We consider the FCN without pre-training and the Dice loss as the *baseline* (using cross-entropy, same structure in Fig. 2 without deep supervision). Variations of systems are created by adding pre-training (PT), Deep Supervision (DS),

Table 1. Ablation study of the proposed method (mean±std.)

Method	Recall	Precision	F_2 score
Baseline	47.9±25.7	62.5±27.2	48.8±24.6
FCN-PT	51.7±21.2	67.8±19.7	53.4±20.6
FCN-PT-DS-1	52.3±23.1	73.2±17.7	54.1±22.4
FCN-PT-DS-2	64.1±22.2	65.4±10.1	62.8±19.8
FCN-PT-DS-3	68.6±22.3	59.7±11.7	65.2±19.2
FCN-PT-DS-4	68.0±22.2	60.8±9.7	64.7±18.8

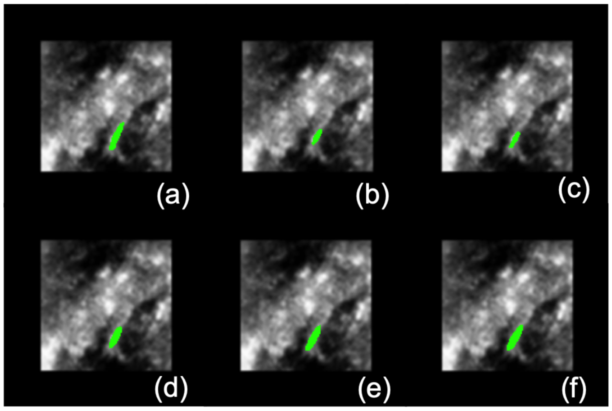

Fig. 3. Examples of ablation study. (a) Annotation, (b) FCN-PT, (c) FCN-PT-DS-1, (d) FCN-PT-DS-2, (e) FCN-PT-DS-3, (f) FCN-PT-DS-4.

which are referred to with their corresponding abbreviations. The weighted F-score loss with deep supervision is noted as *DS*-β, where $\beta = 1, 2, 3, 4$ for Eqn. (4). The results are shown in Table. 1. From the results, several conclusions can be made. First, when compared to the baseline method, pre-trained parameters of the VGG-16 encoding layers provide a better initialization for US segmentation tasks, even the VGG was trained for other tasks in real RGB images other than US data. Second, the deep pyramid supervision is able to improve the segmentation performance in both Recall and Precision. This is because of more discriminative properties are learned at different levels of FCN through multi-scale deep supervision. Third, when β increases, the Recall performance is improved at the cost of Precision, because the weighted loss function in Eqn. (3) emphasizes true positive performance. A high Recall promises a successful segment of the catheter, of which example images are shown in Fig. 3. The visualization shows that a low Recall would lead to a failed segmentation of the catheter in a complex image with anatomy around it, such as in Fig. 3(b) or (c). As observed from the table, the performance saturates with a high β value, which can be explained as the non-linear form of Eqn. (3).

3.2. Performance Comparison with SOTA

We compared FCN-PT-DS-3 with other state-of-the-art (SOTA) medical instrument segmentation methods, i.e., single-scale Gabor-Frangi handcrafted features (GF-SVM) [6], multi-scale-definition handcrafted features for catheter voxels (MS-AdaB) classification [1], triplanar-based voxel-wise classification under CNN (LF-CNN) [10] and shared slice-based FCN fusing prediction from different axes (Share-FCN) [13]. We also considered 3D UNet as suggested in [17], but failed to obtain successful segmentation results.

Table 2. Performance comparison to SOTA (mean±std. for for F_2 score, mean for time estimation)

Method	F_2 score	Time
GF-SVM[6]	2.6±6.2	∼180 sec.
MS-AdaB[1]	34.6±21.3	∼600 sec.
LF-CNN[10]	64.0±16.3	∼110 sec.
Share-FCN[13]	45.9±21.2	∼3.0 sec.
Proposed	65.2±19.2	∼1.1 sec.

The results with corresponding evaluation metrics are shown in Table. 2. As for conventional handcrafted feature methods (GF-SVM and MS-AdaB), the performances are worse than deep learning approaches, due to their limited capacity of feature design and complex anatomical structures in 3D US. For voxel-wise LF-CNN, it holds that despite providing a similar performance than the proposed method, its voxel-wise classification procedure requires a large inference time for each 3D US image of typically over 100 seconds. Instead, the proposed method only spends $\sim$ 1.1 seconds per volume because of its slice-based strategy and 2D operations, leading to 100 times faster inference. As for ShareFCN, it has similar intuition with the proposal, but it is fusing the prediction scores through different principal axes, which leads to an under-segmentation of a catheter with low Recall performance. Moreover, it does not consider the deep supervision scheme that its capacity of semantic information is not fully exploited. For 3D UNet, which is designed for 3D US segmentation in [17], it fails to successfully segment the catheter in our challenging US images. This can be explained as over-fitting using suggested input patches of size 64^3 voxels, clearly requiring more experiments.

4. CONCLUSIONS

In this paper, we propose an efficient FCN-based catheter segmentation in 3D US data, which achieves state-of-the-art performance. By employing a deep supervision scheme with F-score like loss function, our method can exploit the semantic information at different scales, while balancing the trade-off between false positive and false negative samples. With experiments on our challenging *ex-vivo* dataset, our method achieves 68.6% Recall and 59.7% Precision with an efficient inference time of around 1 second. However, slice-based 2D FCN lacks 3D contextual information along the axial direction, thereby limiting the performance. Therefore, further improvement of the scheme is necessary in the future.

5. REFERENCES

[1] H. Yang, C. Shan, A. Pourtaherian, A. F. Kolen, and P. H. N. de With, "Catheter segmentation in three-dimensional ultrasound images by feature fusion and model fitting," *Journal of Medical Imaging*, vol. 6, no. 1, pp. 015001, 2019.

[2] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 130–137.

[3] K. Cao, D. Mills, and K. A. Patwardhan, "Automated catheter detection in volumetric ultrasound," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. IEEE, 2013, pp. 37–40.

[4] M. Uherčík, J. Kybic, Y. Zhao, C. Cachard, and H. Liebgott, "Line filtering for surgical tool localization in 3d ultrasound images," *Computers in biology and medicine*, vol. 43, no. 12, pp. 2036–2045, 2013.

[5] Y. Zhao, C. Cachard, and H. Liebgott, "Automatic needle detection and tracking in 3d ultrasound using an roi-based ransac and kalman method," *Ultrasonic imaging*, vol. 35, no. 4, pp. 283–306, 2013.

[6] A. Pourtaherian, H. J. Scholten, L. Kusters, S. Zinger, N. Mihajlovic, A. F. Kolen, F. Zuo, G. C. Ng, H. H. M. Korsten, and P. H. N. de With, "Medical instrument detection in 3-dimensional ultrasound data volumes," *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1664–1675, 2017.

[7] H. Yang, C. Shan, A. Pourtaherian, A. F. Kolen, and P. H. N. de With, "Feature study on catheter detection in three-dimensional ultrasound," in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. International Society for Optics and Photonics, 2018, vol. 10576, p. 105760V.

[8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[9] A. Pourtaherian, F. G. Zanjani, S. Zinger, N. Mihajlovic, G. C. Ng, H. M. Korsten, et al., "Improving needle detection in 3d ultrasound using orthogonal-plane convolutional networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 610–618.

[10] H. Yang, C. Shan, A. F. Kolen, and P. H. N. de With, "Catheter detection in 3d ultrasound using triplanar-based convolutional neural networks," in *International Conference on Image Processing (ICIP) , In 2018 25th IEEE International Conference on*. IEEE, 2018, pp. 371–375.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[13] A. Pourtaherian, F. G. Zanjani, S. Zinger, N. Mihajlovic, G. C. Ng, H. M. Korsten, et al., "Robust and semantic needle detection in 3d ultrasound using orthogonal-plane convolutional neural networks," *International journal of computer assisted radiology and surgery*, pp. 1–13, 2018.

[14] L. Wang, C. Lee, Z. Tu, and S. Lazebnik, "Training deeper convolutional networks with deep supervision," *arXiv preprint arXiv:1505.02496*, 2015.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer, 2017.

[17] X. Yang, L. Yu, S. Li, X. Wang, Na. Wang, J. Qin, D. Ni, and P. Heng, "Towards automatic semantic segmentation in volumetric ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 711–719.