AUTOMATED CATHETER LOCALIZATION IN VOLUMETRIC ULTRASOUND USING 3D PATCH-WISE U-NET WITH FOCAL 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

3D ultrasound (US) imaging has become an attractive option for image-guided interventions. Fast and accurate catheter localization in 3D cardiac US can improve the outcome and efficiency of the cardiac interventions. In this paper, we propose a catheter localization method for 3D cardiac US using the patch-wise semantic segmentation with model fitting. Our 3D U-Net is trained with the focal loss of cross-entropy, which makes the network to focus more on samples that are difficult to classify. Moreover, we adopt a dense sampling strategy to overcome the extremely imbalanced catheter occupation in the 3D US data. Extensive experiments on our challenging *ex-vivo* dataset show that the proposed method achieves an F-1 score of 65.1% for catheter segmentation, outperforming the state-of-the-art methods. With this, our method can localize RF-ablation catheters with an average error of 1.28 mm.

Index Terms— 3D ultrasound, Catheter localization, 3D U-Net, Dense sampling, Focal loss.

1. INTRODUCTION

Cardiac catheterization has been extensively applied during interventional therapy for heart diseases, which provides a shorter recovery period and lower risk for the patients. Because a direct view on the organ is occluded by skin or other tissues during the operation, medical imaging methods, such as fluoroscopic (X-ray) or ultrasound (US) imaging, have been used to guide the instruments. The maturity of 3D cardiac US imaging, e.g., Transesophageal Echography (TEE), makes 3D US-guided operations an attractive option. 3D US imaging provides richer spatial information of the tissue than the traditional 2D X-ray imaging; more crucially, it is radiation-free. Nevertheless, the low resolution and low contrast of 3D US make it difficult for the physician or sonographer to localize the catheter during surgery in a timely manner. For this reason, computer-aided catheter localization in the 3D US can be helpful.

As a promising technique, image-based catheter localization has been studied in recent years. Frangi features combined with supervised machine learning were investigated for instrument localization in US [1]. In another work, Gabor features combined with Frangi features were studied for catheter localization in the phantom heart data [2]. More recently in [3], more discriminative features were introduced for catheter localization using *ex-vivo* data. In recent years, deep learning methods, e.g., convolutional neural networks (CNNs), have showed significant performance improvement in medical image analysis, including instrument localization in 3D US [4][5]. However, the existing voxel-wise 2.5D processing in [4][5] fails to consider the 3D structure information. Alternatively, a direction-fused fully convolutional network (FCN) was proposed to exploit semi-3D information for catheter segmentation in 3D US [6]. Nevertheless, this 2.5D approach still has a limited capacity for exploiting the 3D context.

Semantic segmentation methods are usually directly applied to the whole 2D image. However, for 3D volumetric data, because of the limited training images and the large size of 3D images [7], which cannot fit the GPU, the 3D medical images are commonly processed by a patch-wise strategy [8][9]. As mentioned in the literature[8][9], the patches for training network are extracted or cropped randomly to follow the original information distribution. Nevertheless, this approach does not work for training a catheter segmentation network in the 3D US, because of the very small occupation of the catheter in 3D US (typically less than 1% of the full image). Moreover, the noisy and low-contrast US images make it even difficult to achieve catheter segmentation specially at the boundary.

In this work, we propose a method for catheter localization in 3D cardiac US using the patch-wise 3D U-Net with model fitting. To address the above difficulties, we adopt a dense sampling strategy on training data, to enhance the catheter information in 3D US and allow the network to concentrate on the relevant information. For 3D U-Net, we consider a focal loss with the adaptive class weight, which enforces the network to focus more on the voxels that are difficult to classify. Based on the segmented volume from 3D U-Net, the catheter is further localized using a RANSAC-based model fitting algorithm [10]. Our experiments on the *ex-vivo* dataset show that the proposed method achieves an F-1 score of 65.1% for catheter segmentation, outperforming the state-of-the-art methods. Our method can localize RF-

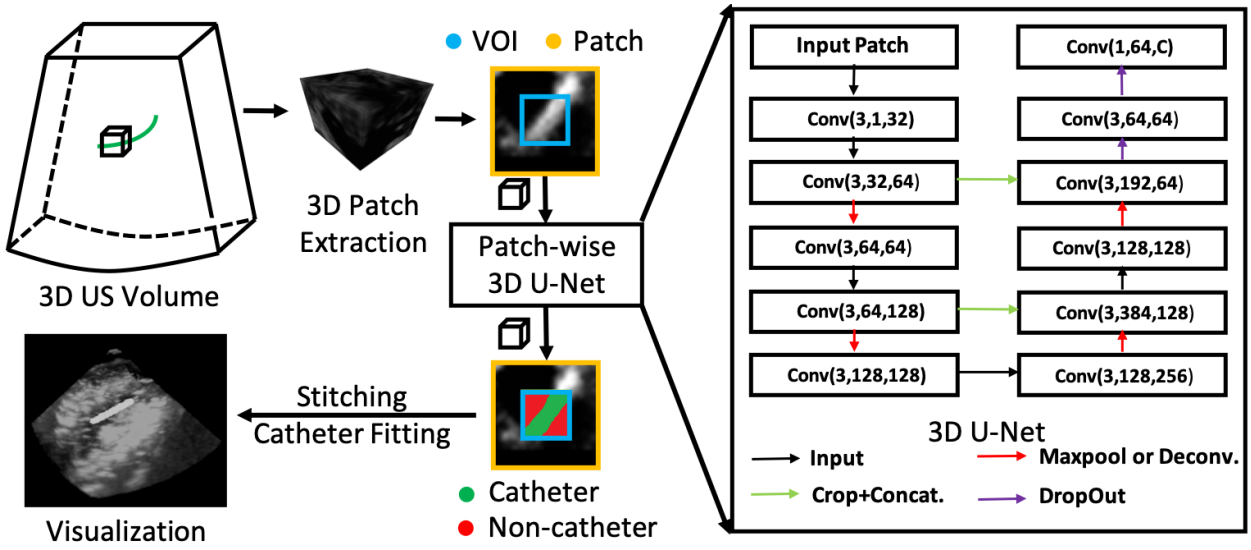

Fig. 1. Block diagram of the proposed catheter localization method based on patch-wise 3D U-Net.

ablation catheters with an average error of 1.28 mm.

2. METHODS

The proposed catheter localization method is shown in Fig. 1. The input volume is first decomposed into small patches. The 3D patch is then segmented by the 3D U-Net, which leads to a smaller VOI output. The segmented VOIs are stitched as the full 3D US volume. Finally, catheter model-fitting is applied to localize the catheter in the noisy 3D segmentation domain. More details are discussed in the following sections.

2.1. Dense Sampling for Training

The U-Net is normally applied to the full 2D image. However, for 3D volumetric data, due to the large size of 3D images and limited GPU memory, patch-wise processing is commonly used [8][9]. Nevertheless, the way of how to extract the patch from training images can influence the performance of the network. It was suggested in [8][9] to crop the patches from the training images randomly. But we found this strategy is not suitable for catheter segmentation in 3D US, because of the very small occupation of the catheter in 3D US (typically less than 1% of the full image). To address this, we propose a dense sampling strategy. More specifically, for each catheter voxel in the training data, a surrounding 3D patch is extracted with the catheter voxel as the center of the patch. Moreover, we randomly downsample the non-catheter voxels to the same amount of catheter voxels for training. This strategy can enhance the catheter information observed by the network and can enforce the network to focus on relevant information. As depicted in Fig. 1, the output (with the size of $N \times N \times N$ voxels) from 3D U-Net is smaller than the input patch (with the size of $M \times M \times M$ voxels).

2.2. 3D U-Net with Focal Loss

We re-design a 3D U-Net as the trade-off between the complexity of the US image and the catheter size. The U-Net has three levels of processing, going from local feature extraction to contextual feature extraction (lies in the horizontal direction). The detailed structure is shown in the right part of Fig. 1. The operations Conv(k, q, p) represent the 3D convolutional operations with kernel size $k \times k \times k$ voxels, applied to the patches with input feature maps q and output feature maps p. After the convolution, the Rectified Linear Unit (ReLU) [12] and Batch Normalization (BN) [13] are applied to the feature maps, which accelerate the convergence speed. The black arrows indicate the flow direction of the feature maps, the green arrows represent cropping and concatenating to combine the feature maps from different feature levels, the red arrows represent Maxpooling (left side) or Deconvolution (right side) upsampling operations by a factor of 2 between different levels [7], and finally, the purple arrows introduce an extra Dropout layer to avoid overfitting [14]. The parameter C represents the number of classes, which is 2 in this paper.

The network is learned by minimizing the focal loss [15], which is shown in Eqn. 1. y_c is the one-hot label of the voxel, $\hat{y}_c$ is corresponding prediction probability of the voxel, w_c is inverse value of ratio of voxel class of the mini-batch in our paper, and γ is the focal parameter to control the ability of concentration, which is selected as 2 in this paper [15]. The w_c can automatically re-balance the class weight for catheter and non-catheter, which is manually selected as a hyper-parameter in [15]. The term γ can control the importance

of voxels that is hard to classify, which push the network to concern more on mis-classified samples.

$$L(y, \hat{y}) = -\sum_{c} w_c \cdot (1 - \hat{y}_c)^{\gamma} \cdot \log(\hat{y}_c). \quad (1)$$

For each iteration in training, the data augmentation methods such as mirror, rotation, contrast and brightness transformation are performed randomly on-the-fly on 3D patches to generalize the network. The parameters of the 3D U-Net are learned by using the Adam optimizer [16] with an initial learning rate of 0.001. The dropout probability is 0.5 in training. The mini-batch size is 32.

2.3. Catheter Model Fitting

The patch is predicted by 3D U-Net to generate the segmented VOI, which is stacked to generate the segmented volume. The segmented volume includes some false positives. We use Sparse-plus-dense (SPD) RANSAC [10][11] algorithm to localize the catheter in noisy segmentation. In the classified binary image after segmentation, named dense volume, the voxels are clustered through connectivity analysis. For each cluster, its skeleton is extracted along the dominant direction of that cluster. These skeletons generate a sparse volume. When fitting the catheter model, three control points are randomly selected and re-ranked by direction analysis from the sparse volume. Then a cubic spline is applied to generate the catheter model. The model of the dense volume that includes the highest number of dense voxels, is chosen as a catheter. In this paper, the threshold controlling the ratio of inliers and outliers is chosen as three voxels. The SPD approach avoids redundancy in the selection of boundary points during RANSAC iterations.

3. EXPERIMENTAL RESULTS

The proposed method was evaluated on a 3D US dataset of porcine heart, which was collected by a 2-7 MHz phase array transducer (TEE), when inserting an RF-ablation catheter (Boston Scientific, diameter 2.3 mm) into the heart chamber. The dataset consists of 25 images, which are individually extracted from 25 separate 3-second videos. The original voxels were re-sampled to obtain an isotropic voxel size in each direction (volume size: $128 \times 128 \times 128$ voxels, space unit: 0.54 mm/direction). The volumes were manually annotated as two classes, catheter and non-catheter, by clinical experts. During our experiments, 3-fold cross-validation was applied on the *ex-vivo* dataset.

The segmentation performance was evaluated using the metrics Recall, Precision, and F1-score. As for U-Net, we first tested three different VOI sizes with standard *cross-entropy loss*: $8 \times 8 \times 8$ (U-Net-8), $16 \times 16 \times 16$ (U-Net-16) and $32 \times 32 \times 32$ (U-Net-32), which lead to the following patch sizes $40 \times 40 \times 40$, $48 \times 48 \times 48$ and $64 \times 64 \times 64$ voxels. The results are shown in Table. 1, with some example results of catheter segmentation shown in Fig. 2. From the results, apparently the optimal VOI size is 16, since the U-Net-16 performs better than U-Net-8 and U-Net-32. Comparing to U-Net-8, the U-Net-16 shows the benefit of a larger perception field that can capture richer spatial information. In contrast, a VOI with 8 voxels only includes a relatively small space when compared to a catheter diameter (of around 5 voxels), which complicates the segmentation. The performance of U-Net-32 is slightly worse than U-Net-16. This can be explained by the catheter size in the image. When the patch size increases, the catheter occupies a smaller part of the patch, and it becomes more difficult to segment the catheter voxels precisely.

Table 1. Average performance of catheter segmentation with different VOI sizes (mean$\pm$ std.).

Method	Recall	Precision	F_1 score
U-Net-8	57.4 ± 14.7	55.9 ± 8.9	56.0 ± 10.7
U-Net-16	59.5 ± 12.9	69.3 ± 10.2	63.1 ± 9.3
U-Net-32	56.6 ± 15.5	68.5 ± 9.7	60.9 ± 11.9

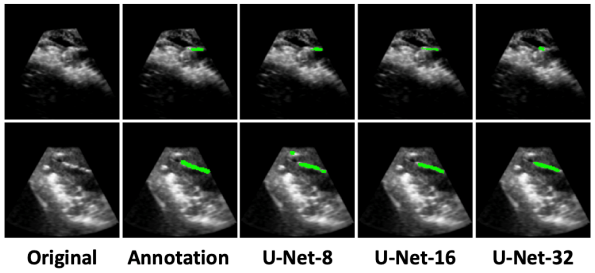

Fig. 2. Examples of catheter segmentation with different VOI sizes. Top to bottom: slices of different volumes, tuned to achieve the best view. Left to right: Original image, manual annotation, U-Net-8, U-Net-16 and U-Net-32.

For U-Net-16, we perform an ablation study for dense sampling and focal loss (all the cases were trained with the same number of iterations): (1) using the random sampling strategy suggested by [8][9], which is noted as U-Net-RS; (2) our proposed dense sampling without sampling on non-catheter voxels, which is noted as U-Net-w/o; (3) our proposed dense sampling, which is noted as U-Net-w; (4) and the focal-loss cross entropy which is applied on U-Net-16, noted as U-Net-FL. The results are shown in Table. 2. The ablation

Table 2. Ablation study with different settings (mean$\pm$ std.).

Method	Recall	Precision	F_1 score
U-Net-RS	53.4$\pm$24.1	42.8$\pm$18.3	44.5$\pm$16.7
U-Net-w/o	63.9$\pm$11.1	15.6$\pm$6.2	24.6$\pm$8.4
U-Net-w	59.5$\pm$12.9	69.3$\pm$10.2	63.1$\pm$9.3
U-Net-FL	62.1$\pm$13.0	70.0$\pm$8.4	65.1$\pm$9.1

Table 3. Performance comparison with the state-of-the-art methods (% with mean± std.).

Method	Recall	Precision	F_1 score
MS-AdaB [3]	45.5±23.2	29.4±11.5	29.4±15.3
ShareCNN[4]	41.4±14.6	55.9±10.9	45.9±11.3
LateCNN [5]	42.8±14.1	70.1±12.2	51.6±11.4
DF-FCN [6]	50.1±17.2	72.5±11.6	57.7±12.0
Proposed	62.1±13.0	70.0±8.4	65.1±9.1

study shows the proposed dense sampling can make the network learn the semantic information successfully while the random selection [8][9] have a worse performance. This is because our sampling method changes the catheter distribution in observation space that network only focuses on a concentrated space while the random sampling approach would push the network focus on the non-catheter area due to the relatively small size of a catheter in US images. Comparing U-Net-w/o to U-Net-w, the downsampling on non-catheter voxels introduces extra-anatomical information of 3D US, which allows the network to learn sufficient information irrelevant to the catheter and improve the performance of Precision. Furthermore, the focal loss in Eqn. 1 can let the network to concentrate on the difficult voxels in the boundary, which is the reason why U-Net-FL achieves a higher Recall performance.

Our proposed method was compared to the state-of-the-art method using multi-scale and multi-defination handcrafted features with an AdaBoost classifier (MS-AdaB) [3], a tri-planar based approach for voxel-wise classification (ShareCNN for needle segmentation) [4], our previously proposed late-fusion CNN for catheter detection (LateCNN, not considering weighted cross-entropy) [5] and direction-fused FCN for catheter segmentation in 3D US (DF-FCN) [6]. The results are shown in Table. 3. The proposed patch-wise 3D U-Net outperforms the handcrafted features based method [3], the tri-planar CNN method [4][5] and 2.5D FCN with transfer learning [6]. More specifically, because of limited discriminating capacity of handcrafted features, they cannot accurately extract meaningful information from noisy and low contrast US images. As a result, more efforts are still needed to achieve a satisfied performance for conventional machine learning methods. As for voxel-wise CNN classification methods, they consider a tri-planar approach that 3D information are degrading during the processing. Moreover, the lack of contextual information is also the reason that [4][5] receives lower performance when compared to our method. Direction-fused FCN is also worse than patch-wise 3D U-Net. Although it makes use of pre-trained VGG network and fuses 3D information by matrix operation at feature space, the limited 3D capacity and huge amount of training parameters limit its performance on our noisy and low contrast dataset, thus leads to a worse performance on Recall.

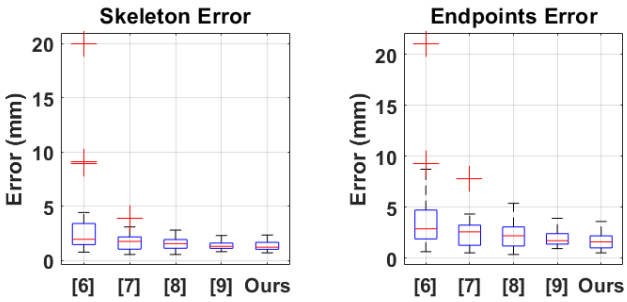

Fig. 3. Localization error distributions for different methods after the model fitting. Handcrafted method [6], ShareCNN [7], LateCNN [8], DF-FCN [9] and Ours.

Last, the catheter localization is achieved by modeling fitting. The accuracy of localization was measured by the skeleton error and the position errors of the two endpoints (which is averaged as 'endpoints' error). The skeleton error is the average distance between 5 equally-sampled points on the fitted skeleton and the annotation skeleton. The endpoints error is the average distance between the endpoints on the localized catheter and the corresponding points on the annotation. The error distributions are shown in Fig. 3 using box plots, which shows our method achieves a higher accuracy because of a higher segmentation performance. From the observation, a higher Recall performance would lead to a more accurate localization in both skeleton error and end-point error. The average skeleton error is 1.28 mm, which is better than the endpoints error (1.60 mm). This difference is explained by a weakness of the RANSAC-based fitting method: it introduces random selection among the skeleton, so that correct endpoints cannot be always be localizeded. In the contrast, the position error, i.e., skeleton error, of whole catheter body can be localized more accurately. In our dataset, catheters can be localized with a skeleton error less than the catheter diameter, which shows a promising performance for interventional guidance.

4. DISCUSSIONS AND CONCLUSIONS

We have presented a catheter localization method for 3D US images, which is based on patch-wise 3D U-Net and a model-fitting algorithm. As verified on the *ex-vivo* dataset, the proposed dense sampling and focal loss show a significant improvement when compared to the state-of-the-art methods (F1-scores achieved 65.1%, $p < 0.05$ in *t-test*). The results show that our method can localize a catheter with an error less than the catheter diameter in 3D US with complex anatomical structures. In future work we will evaluate the method on a larger clinical dataset and improve its computational efficiency.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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 *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 371–375.

[6] H Yang, C Shan, Alexander F Kolen, et al., "Improving catheter segmentation & location in 3d cardiac ultrasound using direction-fused fcn," in *International Symposium on Biomedical Imaging 2019*, 2019.

[7] Ö. ÇIçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[8] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[9] 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.

[10] C. Papalazarou, P. H. N. de With, and P. Rongen, "Sparse-plus-dense-ransac for estimation of multiple complex curvilinear models in 2d and 3d," *Pattern Recognition*, vol. 46, no. 3, pp. 925–935, 2013.

[11] 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.

[12] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[13] S. Loffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR.–2015.–Vol. abs/1502.03167.–URL: http://arxiv. org/abs/1502.03167*, 2015.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[15] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.