CATHETER DETECTION IN 3D ULTRASOUND USING TRIPLANAR-BASED CONVOLUTIONAL NEURAL NETWORKS

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) image-based catheter detection can potentially decrease the cost on extra equipment and training. Meanwhile, accurate catheter detection enables to decrease the operation duration and improves its outcome. In this paper, we propose a catheter detection method based on convolutional neural networks (CNNs) in 3D US. Voxels in US images are classified as catheter (or not) using triplanar-based CNNs. Our proposed CNN employs two-stage training with weighted loss function, which can cope with highly imbalanced training data and improves classification accuracy. When compared to state-of-the-art handcrafted features on *ex-vivo* datasets, our proposed method improves the F2-score with at least 31%. Based on classified volumes, the catheters are localized with an average position error of smaller than 3 voxels in the examined datasets, indicating that catheters are always detected in noisy and low-resolution images.

Index Terms— 3D ultrasound, catheter detection, convolutional neural network, catheter model fitting.

1. INTRODUCTION

In the past years, interventional therapy, such as cardiac catheterization, has been widely applied to achieve lower risk and shorter recovery time for the patient. During the interventional procedures, fluoroscopic imaging or ultrasound (US) are used to guide the catheter or other devices. With the introduction of cardiac 3D US imaging, such as Transesophageal Echography (TEE), the US guided interventions become attractive due to less radiation exposure and richer 3D information. However, 3D US suffers from low signal-to-noise ratio and low spatial resolution, which makes the sonographer or physician to spend more time to localize the catheter during the operation. Therefore, automatically detecting the catheter inside the US image can be helpful during the operations.

Several approaches have been proposed during the past years, like robotic-based catheter detection or electrical-magnetic sensor based tracking. Nevertheless, these approaches require extra equipment or training, which introduce more cost when employing them into clinical application. Alternatively, image-based catheter detection is attractive without introducing extra devices. An US image-based catheter detection method was proposed using template matching [1]. However, it requires prior knowledge of catheter direction and lacks discriminative information, so that it fails in complex anatomical environments and variable catheter poses. Alternatively, the image intensity together with Frangi filter response were used as discriminative features for the supervised learning approach, which showed promising results in medical tools detection [2]. An in-depth study [3] on supervised instrument-voxel classification showed that employing the tool shape description and its intensity achieved better results than Frangi-based method [2]. However, this filter-bank based approach was only verified on an *in-vitro* phantom dataset, while its performance on complex *ex-vivo* or *in-vivo* tissue still needs to be tested. In our previous study [4], the multiscale-based features on *in-vitro* and *ex-vivo* datasets, which showed a significant performance improvement on catheter detection. However, robust detection of the catheter is important for interventions in complex anatomical environments, requiring a further improvement on catheter detection.

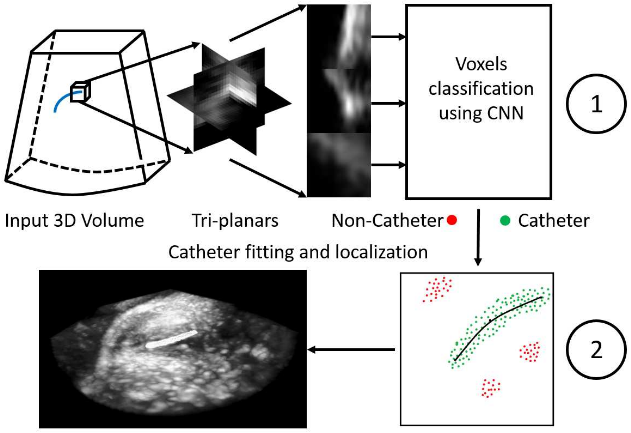

Voxels classification using CNN
1
Input 3D Volume
Tri-planars
Non-Catheter ●
● Catheter
Catheter fitting and localization
2

Fig. 1. The proposed catheter detection system.

Recently, deep learning, e.g. convolutional neural networks, which have shown significant improvement in medical image-based applications [5]. In this paper, a CNN-based catheter-voxel classification method is proposed to achieve a high recall while keeping a high precision. The experiments on *ex-vivo* datasets lead to more robust catheter detection under model fitting [6].

2. METHODS

The proposed catheter detection system in Figure 1, which includes two processing steps. The first stage is catheter-like voxel classification using three orthogonal slices of the 3D input volume. The second stage consists of localizing the catheter via pre-defined models and highlight it for 3D rendering or X-plane visualization.

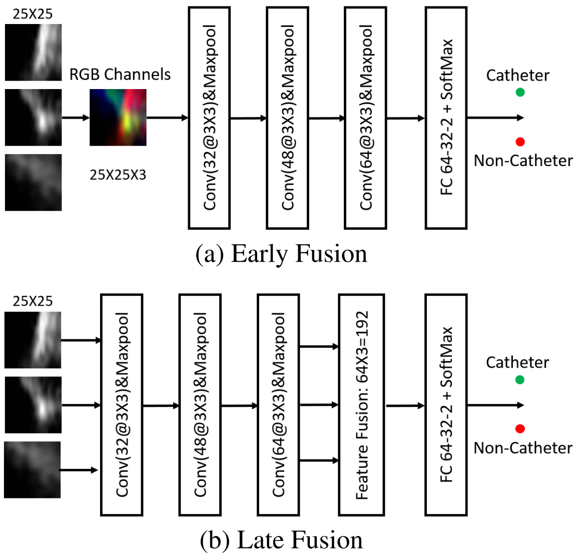
2.1. Stage 1: Catheter-like voxel classification

In order to robustly classify catheter voxels from other anatomical structures, like heart valve or heart wall, we employ a CNN for classification. This method gives each voxel a binary label based on the voxel's local information from the input 3D US volume. The neighborhood information can be extracted in various ways, by a) 2D slices without considering spatial dependency [7], b) using a 2.5D image with spatial dependency [8], c) employing a true 3D cube [9]. In this work, we consider 2.5D-based networks in more detail, thereby balancing the trade-off between computational complexity and classification accuracy. For each voxel to be classified, a reference cube is constructed by considering it around a center point. After the cube is extracted, three orthogonal planes passing through the center point are extracted as the input for the network. In this work we extract a cube with a size of $25 \times 25 \times 25$ voxels to capture sufficient catheter shape information.

2.1.1. *Network Architectures of the CNN*

To make use of spatial information from 3 orthogonal planes, we employ two different networks, which fuse spatial information at an early stage or a late stage. We named them early fusion (E-CNN) and late fusion (L-CNN), respectively [10], of which the architectures are shown in Figure 2. The E-CNN treats three different directions as different color channels, which is similar to the one used in CT image detection [8]. The L-CNN tries to train a shared network (i.e. parameters and operations) on different input slices [11].

Both architectures have similar layers, except the first layer of the convolutional part, which are shown in Figure 2. For the E-CNN case, the filter kernel size is $3 \times 3 \times 3$ elements in the first convolutional layer, while the L-CNN has $3 \times 3 \times 1$ elements. The early CNN processes the spatial information during the first CNN layer, while the other network combines the spatial information prior to the fully connected (FC) layers by concatenating feature maps from the CNN output. Besides three convolutional layers, three fully connected layers with 64, 32 and 2 neurons are used prior to the softmax layer. For these architectures, the trainable parameters are around 49,000 and 56,000, respectively.



(a) Early Fusion

(b) Late Fusion

Fig. 2. CNN Architectures.

2.1.2. *Training Stage*

During the training, one of the most challenging problems in our case is the number of catheter voxels being much smaller than the amount of non-catheter voxels, leading to a ratio of 1/2,500. In order to avoid bias from the imbalanced training data, we employ a two-stage training method. First, the imbalanced training samples are re-sampled on non-catheter voxels to obtain the same size as catheter voxels. The network is trained on the re-sampled data. Training volumes are then validated by this model, where the falsely classified voxels are selected as additional training data to update the network [11]. This updating method fully exploits imbalanced information without focusing on easily classified samples. In order to improve the orientation-invariant property, the 3D cube is randomly mirrored in one of eight different possibilities prior to extracting the three slices.

The parameters are learned by minimizing the cross-entropy loss function using Stochastic Gradient Descent with Momentu [12]. During the training stages, the loss functions are characterized in different forms. The updating method tends to achieve a higher precision value, while sacrificing the recall if the amount of false positive samples is larger than the true positives. In our problem, due to the complex anatomical structure inside the heart, the number of false positives is usually several times larger than the number of catheter-like voxels. To preserve a high recall, the loss function is redefined as weighted cross-entropy to balance the training sample ratio in second training, see Eqn. (1). The y indicates the label of the sample while $\hat{p}$ is the class probability of the sample, and parameter w is the sample class ratio among the training mini-batch. The loss is now specified as

$$Loss(y, \hat{p}) = -(1-w)ylog(\hat{p}) - w(1-y)log(1-\hat{p}). \quad (1)$$

The ReLU [13] is applied to introduce the non-linearity. To prevent overfitting, the dropout [14] is used with a probability of 50% in FC layers during the training, together with an L2 regularization with 0.0001 regulation strength. Initial learning rates are 0.01 and 0.001, for the re-sampled stage and updating stage, respectively.

2.2. Stage 2: Catheter Localization

The classified volume includes outliers, which have a catheter-like structure. To robustly localize the catheter, we employ the RANSAC algorithm [15] to fit a predefined catheter model. The catheter is modeled by a curved cylinder with a fix radius, which is set to 3 voxels in our experiments. In order to fit the curved model robustly, we modified the RANSAC algorithm based on [6]. As for a classified volume, so-called dense volume, the positively classified voxels are clustered by connectivity analysis. The centerlines of clusters are extracted by finding the centerpoint of each cross-section along the most dominant direction. These centerlines are called sparse volume (when the largest cluster size >1000, otherwise a sparse volume equals to a dense volume). During the fitting stage, three points are randomly sampled from the sparse volume, then a direction analysis is applied to find their ranking direction order. The order ensures that the cubic spline is fitting to pass the points in sequential order, which generates the skeleton of the catheter model. The model of the dense volume that includes the highest number of catheter voxels is chosen as a catheter.

3. EXPERIMENTAL RESULTS

Our method was evaluated on 3D US datasets of porcine hearts, which were acquired by a 2-7 MHz phase array in Datasets 1 and 2 and a 1-5 MHz phase array in Dataset 3. The voxels were re-sampled to obtain equal voxel size in each direction. During the experiment, 4-fold cross-validation was applied on all 32 volumes to separate training and testing data.

3.1. Classification Performance

During the testing, the classification performances of networks were evaluated by recall (R), precision (P) and F2-score, as shown in Eqn. (2). These metrics were used due to imbalanced data in testing volume and we were more interested in catheter-like voxels. Besides these metrics, our CNN methods were compared to the state-of-the-art handcrafted features [3] and [4]. In [3] and [4], the performance values were obtained by tuning the thresholds to achieve the highest F-2 score in each fold, computed by

Table 1. Characterization of datasets

Data Set	Vol. #	Volume Size	Voxel Size
Dataset 1 (*ex-vivo*)	10	$179 \times 175 \times 92$	0.4 mm
Dataset 2 (*ex-vivo*)	10	$102 \times 69 \times 92 \sim$ $193 \times 284 \times 190$	0.6 mm
Dataset 3 (*ex-vivo*)	12	$137 \times 130 \times 122$	0.7 mm

Table 2. Average performance of voxel-based classification

Method	Recall	Precision	F2-score
[3]	0.34±0.28	0.14±0.16	0.20±0.15
[4]	0.69±0.17	0.31±0.06	0.54±0.11
E-CNN	0.75±0.22	0.60±0.10	0.70±0.18
L-CNN	0.76±0.22	0.59±0.10	0.71±0.18

$$F_2 = \frac{5 \cdot P \cdot R}{4 \cdot P + R}. \quad (2)$$

Table 2 demonstrates that when compared to [3][4], the CNN-based approaches have at least 31% improvement in F2-scores, while the recall and precision are improved at least by 10% and 90%, respectively. Figure 3 depicts the recall value distributions for 4 different networks, i.e. E-CNN / L-CNN with weighted loss functions and E-CNN* / L-CNN* without weighted loss function. From the distributions without redefined loss function, the networks tend to obtain lower recall values, sometimes an extremely low recall is achieved around 0.08 in the fourth column. When comparing E-CNN to L-CNN, the latter one has a better recall distribution with a higher median value and less outliers, which can be explained by more complete fusing of spatial information after the convolutional layers. However, these networks have almost the same complexity, so that the performance difference is small in our limited dataset.

Figure 4 illustrates some classification results. The false positives have similar structure as the catheter-like points, such as heart valves. From the false-negative images, they are mostly located at the catheter edges, or they are too close to the tissue so that the network tends to regard them as negative samples. Figure 5 depicts cross-section images along the axial direction. The classified voxels are annotated inside the red boxes. The examples at the top are selected from the worst results. Due to blurred edges and weak image intensities of the catheter, the network only classifies the most confident points, while ignoring some catheter voxels around them. This may be caused by the low resolution of the ultrasound image where a complex network tends to misclassify catheter

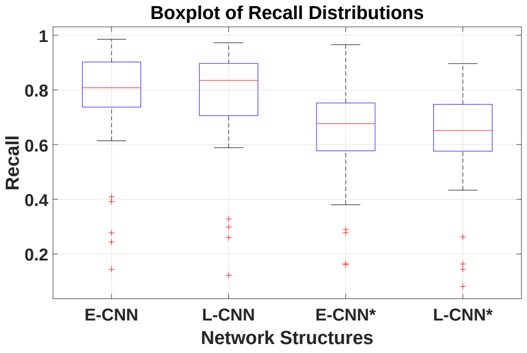


Fig. 3. Recall boxplots for different network structures.

voxels to improve average false positive performance.

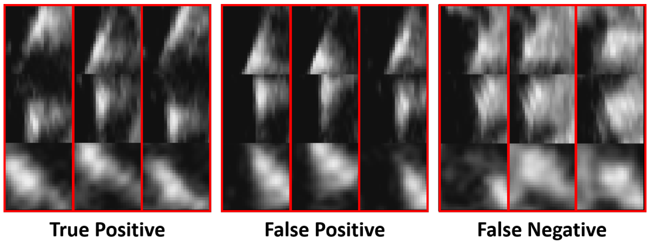


Fig. 4. Examples of different tri-planar classification results. Ordering: lateral (top), azimuthal (middle), axial (bottom).

3.2. Localization Performance

The obtained results with higher accuracy promise a more robust catheter model fitting, or even directly localize the catheter inside the volume. Based on the classified volumes from L-CNN, the catheter model was fitted by the cubic spline-based model. The RANSAC algorithm generated the skeleton of the catheter, which can be used to evaluate the localization accuracy. An example of classified voxels and fitted voxels in the volume are highlighted in Figure 6. To evaluate the detection accuracy, we employed position error and end-point error as metrics. The position error is the average distance of 5 equally-sampled points on the detected skeleton and the ground-truth skeleton. The end-point error is the average distance between two end points on the detected skeleton and the corresponding end points on the ground truth. The performance results are shown in Table 3 for three datasets individually because of the resolution difference (executing the fitting algorithm three times). The average position error on 32 volumes is around 1.7 voxels, which is better than the average end-point error. This shows that the coarse classification-based model-fitting method has worse end-point accuracy due to the blurred-edge phenomena in low-resolution images. But for the localization, the catheter can be detected with a position error smaller than 3 voxels (around the catheter radius).

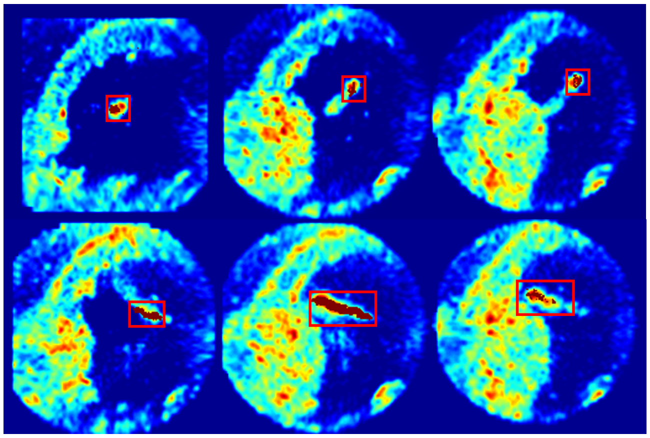


Fig. 5. Classified voxels in two volumes (Top: R=0.12, P=0.36; Bottom: R=0.92, P=0.68). Images are cross-sections along the axial, the intensities are demonstrated by the heat map, while classified voxels are highlighted with red color.

Table 3. Average performance of catheter localization

Dataset	Position Error	End-Point Error
Dataset 1	0.72 mm/1.8 voxel	1.28 mm/3.2 voxel
Dataset 2	0.90 mm/1.5 voxel	1.32 mm/2.2 voxel
Dataset 3	1.12 mm/1.6 voxel	1.47 mm/2.1 voxel

4. CONCLUSION AND DISCUSSION

In this paper, we study novel CNN-based catheter detection in 3D US, which achieves a clearly highier accuracy. By employing CNN with an updating stage combined with a weighted loss function, our proposed method can exploit all information of highly imbalanced datasets, while keeping the trade-off between false positive and false negative samples. When compared to state-of-the-art handcrafted features on challenging *ex-vivo* datasets, our CNN-based method achieves 0.76 recall and 0.59 precision values. This result provides a more robust condition for model fitting. The numerical analysis on catheter localization shows that the proposed system has an average position error smaller than 3 voxels. Therefore, the detected catheter can be easily demonstrated to physicians with automated annotation inside the 3D US images. Further study with larger datasets using different spatial information fusion approaches is needed.

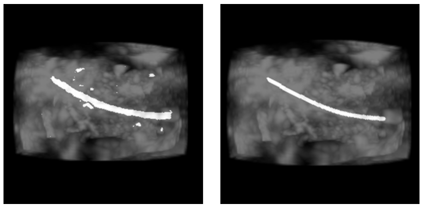


Fig. 6. Left: the classified result; Right: the fitted result.

5. REFERENCES

[1] 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.

[2] 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.

[3] A. Pourtaherian, H. J. Scholten, L. Kusters, S. Zinger, N. Mihajlovic, A. F. Kolen, F. Zou, 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*, 2017.

[4] H. Yang, A. Poutaherian, C. Shan, A. F. Kolen, and P. H. N. de With, "Feature study on catheter detection in three-dimensional ultrasound," in *SPIE Medical Imaging*. International Society for Optics and Photonics, In Press.

[5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak andB. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, 2017.

[6] 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.

[7] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2013, pp. 246–253.

[8] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 520–527.

[9] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 212–220.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. Li, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[11] A. Pourtaherian, F. G. Zanjani, S. Zinger, N. Mihajlovic, G. C. Ng, H. H. M. Korsten, and P. H. N. de With, "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.

[12] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452, 2015.

[13] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

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

[15] M. A. Fischlerand R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.