Deep Q-Network-Driven Catheter Segmentation in 3D US by Hybrid Constrained Semi-Supervised Learning and Dual-UNet

Hongxu Yang[1], Caifeng Shan[2],
Alexander F. Kolen[3], and Peter H. N. de With[1]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
h.yang@tue.nl
[2] Shandong University of Science and Technology, Qingdao, China
[3] Philips Research, Eindhoven, The Netherlands

Abstract. Catheter segmentation in 3D ultrasound is important for computer-assisted cardiac intervention. However, a large amount of labeled images are required to train a successful deep convolutional neural network (CNN) to segment the catheter, which is expensive and time-consuming. In this paper, we propose a novel catheter segmentation approach, which requests fewer annotations than the supervised learning method, but nevertheless achieves better performance. Our scheme considers a deep Q learning as the pre-localization step, which avoids voxel-level annotation and which can efficiently localize the target catheter. With the detected catheter, patch-based Dual-UNet is applied to segment the catheter in 3D volumetric data. To train the Dual-UNet with limited labeled images and leverage information of unlabeled images, we propose a novel semi-supervised scheme, which exploits unlabeled images based on hybrid constraints from predictions. Experiments show the proposed scheme achieves a higher performance than state-of-the-art semi-supervised methods, while it demonstrates that our method is able to learn from large-scale unlabeled images.

Keywords: Catheter segmentation · deep reinforcement learning · Dual-UNet · semi-supervised learning · hybrid constraint.

1 Introduction

Catheter segmentation in 3D ultrasound (US) images is of great importance in computer-assisted intervention, such as RF-ablation or cardiac TAVI procedures [13, 15, 14]. Catheter segmentation in 3D US is usually treated as a voxel-wise classification task, which assigns a semantic category to every 3D voxels (regions) by data-driven methods, such as using Deep Learning (DL) [8]. Typically, DL methods encounter challenges like limited training samples, expensive annotations and imbalanced classes. These factors degrade the DL performances for catheter segmentation in 3D US in different aspects [12]. To overcome the class-imbalance issue, a patch-based strategy was introduced to address the strongly
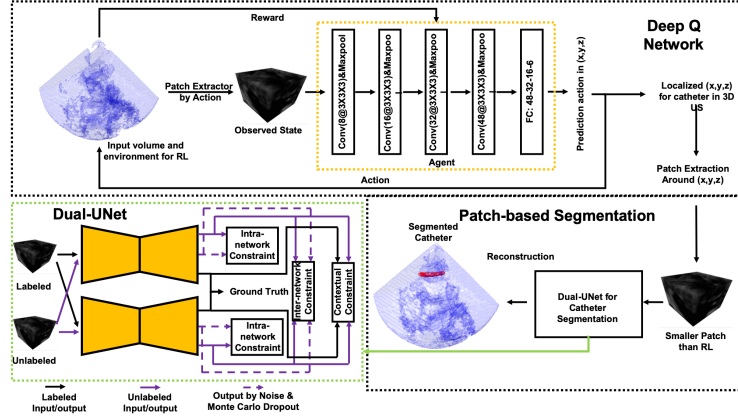
Fig. 1. Schematic view of the proposed framework.

imbalanced class distributions [13, 15]. However, this approach requires an accurate annotation for the training datasets, which is expensive for 3D US images and laborious for clinical experts. Moreover, a patch-based strategy requires the Convolutional Neural Network (CNN) to exhaustively segment the patches on the whole image, which is computationally expensive for real-time clinical usage. In the past years, several methods [17, 9, 7, 2, 10, 16] have been proposed to also utilize the unlabeled images and thereby improve the segmentation performance for medical imaging. Adversarial learning has been studied to make use of unlabeled images, by either enforcing the network to perform segmentation on unlabeled images similar to the labeled images [17], or selecting the most reliable segmentations on unlabeled images to train the segmentation network [9]. Li *et al.* [7] proposed to extend the Π-model for semi-supervised learning, which achieved promising results in skin lesion segmentation. A multi-task-based framework was proposed to exploit unlabeled images by attention-based image reconstruction [2]. An uncertainty-aware self-ensembling model was proposed [10, 16] to make use of certainty estimations for the segmentations of unlabeled images, which enhances the segmentation performance with limited annotations. Although uncertainty-aware methods [10, 16] achieved better performances than state-of-the-art methods (SOTAs), they are all based on the mean-teacher approach with exponential moving averaging on parameter updating, which encounters a parameter-correlation between teacher and student models [4]. As a consequence, the reliability of used voxels may not be stable enough for SSL.

To address above challenges on annotation effort and segmentation strategies for patch-based CNN, we propose a deep Q learning-driven [3] catheter segmentation framework in 3D US by a semi-supervised approach that leverages reinforcement learning as a pre-localization step for catheter segmentation, which makes use of limited labeled and abundant unlabeled images to train a patch-based CNN. The proposed two-stage deep learning scheme is shown in Fig. 1. More specifically, the deep Q network (DQN) method performs coarse catheter

localization by a learned search-path policy through the interaction with the 3D US image, which omits voxel-level annotations. Because of the DQN properties for discrete space actions, the annotation effort is drastically reduced, but the detection is degraded from a full voxel-level annotation to region-level semantic annotation. After the DQN-driven catheter localization, the proposed scheme introduces a Dual-UNet network, which segments the catheter around the localized region by a patch-based strategy [13, 15]. To overcome the limited amount of labeled images and leverage the abundantly available unlabeled images, the proposed Dual-UNet is trained by a novel semi-supervised learning (SSL) scheme. In more detail, the proposed scheme consists of two independent UNets by learning the semantic information from a hybrid loss function, based on a supervised loss, intra-network uncertainty constraint loss, inter-network uncertainty constraint loss, and a contextual constraint loss. With uncertainty estimations within two UNets, the most reliable voxel-level predictions of unlabeled images are used as the semi-supervised signal to enhance the segmentation performance. To compensate the voxel-level constraints from the uncertainty estimation and enhance the usage of semantic information, we also introduce a contextual constraint, which enforces the network to learn the contextual information from unlabeled images. From our experiments, the proposed method was extensively compared to the SOTAs on an *ex-vivo* dataset for cardiac catheterization. The results demonstrate that our method performs better than supervised learning method from same back-bone, while outperforming the SSL SOTAs.

Exploiting the proposed scheme for catheter segmentation, this paper presents the following contributions. We propose a novel catheter segmentation scheme based on the DNQ-driven Dual-UNet, which is trained by the proposed SSL approach. The proposed SSL training strategy is based on hybrid constraints and employs an unlabeled signal to improve the discriminating capacity of the CNN. The method leverages fewer annotations than a supervised learning approach, thereby reducing the training challenges. Moreover, the proposed scheme achieves a higher efficiency because of the use of a coarse-to-fine strategy.

2 Methods

The proposed segmentation method is shown in Fig. 1. First, catheter center is detected by the DQN, which provides an estimation for the catheter location. Second, with the obtained catheter center, Dual-UNet is applied on local patches around the estimated location, which is trained by proposed SSL scheme.

2.1 Deep Q Learning as a pre-localization step for segmentation

The task of extracting a 3D US region containing the catheter can be modeled under the reinforcement learning framework. The agent with its current observation state s interacts with the environment $\mathcal{E}$, by performing successive actions $a \in \mathcal{A}$ to maximize the expected reward r. In this paper, we define the observation state in Cartesian coordinates as a 3D observation patch with size

of d^3 voxels w.r.t. the patch center point (x, y, z). To interact with $\mathcal{E}$, the action space $\mathcal{A}$ has six elements, which are defined as $\{\pm a_x, \pm a_y, \pm a_z\}_{scale}$ w.r.t. three different axes with a resolution of step *scale*. With the observed state s, the agent makes a decision of an action from $\mathcal{A}$, to iteratively update the location of the 3D patch. After each action, a reward RL system is characterized as $r = sign(D(Pt_g, Pt_{t-1})/scale - D(Pt_g, Pt_t)/scale)$, where D denotes the Euclidean distance between two points, Pt_g is the ground-truth point, Pt_t is the current state, Pt_{t-1} is the previous state and *scale* represents the step scale in $\mathcal{E}$ $[1, 3]$. As a result, reward $r \in \{-1, 0, +1\}$ indicates whether the agent invokes a patch to move or leave it to the target point. With the obtained reward, the optimized action-selection policy can be characterized by learning a state-action value function $Q(s, a)$, which can be approximated by the Deep Q Network (DQN). To train the DQN, the loss function is defined as:

$$\mathcal{L}_{\mathrm{DQN}} = E_{s,r,a,\hat{a} \sim M}[(r + \gamma \cdot \max_{\hat{a}} Q(\hat{a}, \hat{s}; \widetilde{\omega}) - Q(a, s; \omega))^2], \tag{1}$$

where the future reward discount parameter γ is set to be 0.9, $\hat{a}$ and $\hat{s}$ are action and observed state in the next step, respectively. Parameter M is the experience replay to de-correlate the random samples. Parameters ω and $\widetilde{\omega}$ are trainable parameters of the Q networks for current and target network, respectively. The architecture of the adapted Q network is depicted in Fig. 1, where four recent patches are used as the input. Search is terminated after local oscillation.

2.2 Semi-supervised catheter segmentation by Dual-UNet

With coarse localization of the catheter in a 3D US volume, the catheter is then segmented by the proposed patch-based Dual-UNet, which is trained by a hybrid constrained framework. The proposed Dual-UNet structure is motivated by the mean-teacher architecture, which learns the network parameters by updating a student network from a teacher network [10, 16]. Intuitively, this method introduces two networks whose parameters are highly correlated due to the strategy of performing an exponential moving average on the updating process. As a result, the obtained knowledge is biased and may not be discriminative enough [4]. Alternatively, we propose to use two independent networks, to learn the discriminating information by knowledge interaction through uncertainty constraints. Moreover, a contextual-level prediction constraint is introduced to compensate the information usage from voxel-level uncertainty estimations.

With the localized catheter from the DNQ, patches around it are extracted for semantic segmentation (or training). We studied the task of catheter segmentation in 3D US volumetric data, where training patches contain N labeled patches and M unlabeled patches. In this paper, labeled patches are denoted by $\{(x_i, y_i)\}_{i=1}^N$ and unlabeled patches are denoted as $\{x_j\}_{j=1}^M$, where $x \in R^{V^3}$ is the 3D input patch and $y \in \{0, 1\}^{V^3}$ is the corresponding ground truth ($V = 48$ as suggested by [15]). The task of semi-supervised learning is to minimize the

following loss function with inputs of patches and ground truth (if applicable):

$$\min_{\omega_1,\omega_2} (\sum_{i=1}^{N} (L_{sup}(f_1(x_i;\omega_1), y_i) + L_{sup}(f_2(x_i;\omega_2), y_i))$$

$$+ \alpha \sum_{j=1}^{M} (L_{intra}(f_1(x_j;\omega_1), f_1(x_j;\omega_1,\xi_1)) + L_{intra}(f_2(x_j;\omega_2), f_2(x_j;\omega_2,\xi_2))$$

$$+ L_{inter}(f_1(x_j;\omega_1), f_2(x_j;\omega_2), f_1(x_j;\omega_1,\xi_1), f_2(x_j;\omega_2,\xi_2)))$$

$$- \beta \sum_{i,j=1,1}^{N,M} L_c(f_1(x_i;\omega_1), f_2(x_i;\omega_2), f_1(x_j;\omega_1), f_2(x_j;\omega_2))), \tag{2}$$

where ω_1, ω_2 are trainable parameters, L_{sup} denotes supervised loss (e.g. cross-entropy loss + Dice loss), L_{intra} is the uncertainty constraint to measure the consistency between outputs from each individual network $f_q(\cdot)$ under different perturbations, where $q \in \{1, 2\}$. Here, ξ involves the dropout layers and Gaussian random noise of the input. Function L_{inter} is an uncertainty constraint to measure the consistency between the two networks $f_1(\cdot)$ and $f_2(\cdot)$, which makes use of information from two independent networks to enhance the performance. Function L_c is the cross-entropy loss w.r.t. the prediction with a label or without label for Adversarial learning [17]. Coefficient α and β were empirically selected as 0.1 and 0.002 to balance the losses.

Uncertainty Intra-network Constraint L_{intra}: Without annotation for an input patch, the prediction from networks to guide the unsupervised learning may be unreliable. To generate a reliable prediction from history and guide the network to gradually learn from the more reliable prediction, we design an uncertainty constraint for each individual network. Given an input patch, T predictions are generated by a forward pass, based on a Monte Carlo Dropout and input with Gaussian noise [5]. Therefore, the estimated probability map is obtained by the average of T times prediction for an input patch, i.e. $\hat{P}_q$ for network $f_q(\cdot)$. Based on the above probability maps, the uncertainty is obtained as $\hat{U}_q = -\hat{P}_q log(\hat{P}_q)$ and the loss constraint for network $f_q(\cdot)$ is formulated by:

$$L_{intra} = \frac{\sum(\mathcal{I}(\hat{U}_q < \tau_1) \odot ||f_q(x;\omega_q) - \hat{P}_q||)}{\sum \mathcal{I}(\hat{U}_q < \tau_1)}, \tag{3}$$

where $\mathcal{I}$ is a binary indicator function, $\tau_1 = -0.5ln(0.5)$ is a threshold to measure the uncertainty, which selects the most reliable voxels by binary voxel-level multiplication $\odot$. Parameter $f_q(x;\omega_q)$ is the prediction from network $f_q(\cdot)$ with $q \in \{1, 2\}$. By following this approach, the proposed strategy is approximately equal to the mean-teacher method with the history step as unity.

Uncertainty Inter-network Constraint L_{inter}: Besides the above uncertainty constraint for each network, we also propose an uncertainty constraint to measure the consistency between two individual networks with the purpose to constrain the knowledge and avoid bias. The proposed uncertainty constraint

on two networks can enhance the overall stability of the model by comparing the predictions between two networks. From the proposed networks in Fig. 1 and the above definitions, different predictions can be obtained, expressed as P_q and $\hat{P}_q$ for normal prediction and Bayesian prediction. Moreover, corresponding binary predictions are obtained as C_q and $\hat{C}_q$, respectively, which are thresholded by 0.5. This leads to a stable prediction, which is formally written as

$$\mathcal{S}_q = \mathcal{I}(C_q \odot \hat{C}_q) \odot (\mathcal{I}(U_q < \tau_2) \oplus \mathcal{I}(\hat{U}_q < \tau_2)), \tag{4}$$

where $\tau_2 = -0.7ln(0.7)$ is a threshold to select the more reliable voxels for $f_q(\cdot)$. Furthermore, we also define the voxel-level probability distance $D_q = ||f_q(x; \omega_q) - \hat{P}_q||$, which indicates the network-wise consistency. Symbol $\oplus$ stands for a voxel-based logical OR. With the above, the uncertainty inter-network constraint L^1_{inter} (where $L_{inter} = L^1_{inter} + L^2_{inter}$) for $f_1(\cdot)$ is formulated by:

$$L^1_{inter} = \frac{\sum(((\mathcal{S}_1 \odot \mathcal{S}_2 \odot \mathcal{I}(D_1 > D_2)) \oplus (\overline{\mathcal{S}_1 \odot \mathcal{S}_2} \odot \mathcal{S}_2)) \odot ||P_1 - P_2||)}{\sum((\mathcal{S}_1 \odot \mathcal{S}_2 \odot \mathcal{I}(D_1 > D_2)) \oplus (\overline{\mathcal{S}_1 \odot \mathcal{S}_2} \odot \mathcal{S}_2))}, \tag{5}$$

where $|| \cdot ||$ is the probability distance at the voxel level and $\overline{(\cdot)}$ is binary NOT operation. This uncertainty constraint enables the unsupervised signal communication between two individual networks for $f_1(\cdot)$ and vice versa for $f_2(\cdot)$.

Contextual Constraint L_c: The above constraints only consider voxel-level consistency of paired predictions, while ignoring the differences between labeled and unlabeled predictions at the contextual level. To enhance the predictions similarity at contextual level, we also introduce a contextual constraint, which is based on the implementation of adversarial learning. The labeled and unlabeled predictions are classified by an adversarial classifier, which is learned by the loss L_c, i.e. BCE, based on whether it has annotation or not. Meanwhile, L_c is maximized by Eqn. (2) to enhance the contextual similarity between predictions.

3 Experimental Results

Materials and Implementation Details: We collected 88 3D US volumes from 8 porcine hearts. During the data collection, the tissues were placed in water tanks with an RF-ablation catheter (diameter ranging from 2.3 mm to 3.3 mm) inside the chambers. The obtained images were re-sampled to have a volume size of $160 \times 160 \times 160$ voxels with padding applied, which leads to a voxel size ranging from 0.3^3-0.9^3 mm. All the volumes were manually annotated at voxel level. Moreover, the catheter centers were also marked as the target location for DQN. To validate proposed method, 60 volumes were randomly selected as training set and 28 volumes were used as testing images. To train the DQN, 60 volumes with target location were used to learn the path-search policy. To train the Dual-UNet, 10%, 20% and 30% of 60 images were selected as the labeled images, while the remainder were the unlabeled images for SSL.

We implemented our framework in *TensorFlow*, using a standard PC with a TITAN 1080Ti GPU. We trained the DQN with Adam [6] optimizer (learning
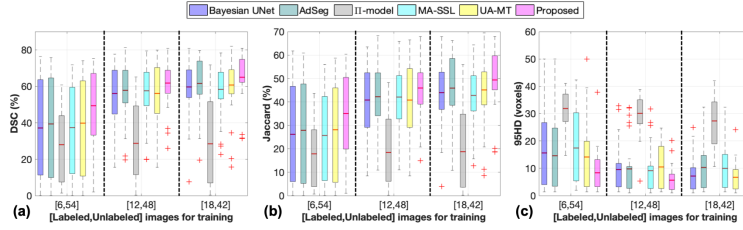
Fig. 2. Comparison to SOTAs by boxplot: (a) Dice Score, (b) Jaccard Score, (c) 95HD. Higher DSC/Jaccard and lower 95HD show better results for segmentation.

Table 1. Segmentation performance by ablation studies in Dice Score (DSC), Jaccard Score (Jaccard) and 95% Hausdorff Distance (95HD), which are shown in mean±std.

Method	# Images Used		Metrics		
	Labeled	Unlabeled	DSC (%)↑	Jaccard (%)↑	95HD (voxels)↓
DU-L_{intra}	6	54	42.4±25.1	29.8±19.2	13.9±11.2
DU-$L_{intra+inter}$	6	54	47.9±23.5	34.2±19.4	11.7±8.3
Proposed	6	54	**49.4±20.7**	**35.1±17.3**	**8.5±7.8**
DU-L_{intra}	12	48	58.8±17.0	43.5±15.9	8.2±9.4
DU-$L_{intra+inter}$	12	48	59.6±13.2	43.7±13.1	7.8±8.5
Proposed	12	48	**61.8±13.4**	**45.9±13.3**	**5.7±4.6**
DU-L_{intra}	18	42	62.6±14.0	46.8±13.4	6.2±6.6
DU-$L_{intra+inter}$	18	42	63.8±13.3	48.1±13.3	5.2±5.9
Proposed	18	42	**65.0±13.5**	**49.4±13.3**	**4.7±5.1**
UNet[15]	60	0	63.0±14.4	47.3±13.2	5.5±7.1

rate 1e-4) for 40 epochs. Replay memory was 1e5. Parameters of the target network were updated for every 2,500 steps. When considering the efficiency and accuracy of the DQN, we defined the input state space to 55^3 voxels for resized images with size 128^3, which promises the observations can contain sufficient catheter contextual information. To construct the Dual-UNet as the segmentation network, we employ Compact-UNet [15] as our backbone, together with a joint cross-entropy and Dice loss. To adapt the UNet as a Bayesian network and generate uncertainty prediction, dropout layers with rate 0.5 were inserted prior to the convolutional layers. Gaussian random noise was also considered during uncertainty estimation. For the uncertainty estimation, $T = 8$ was used to balance the efficiency and quality of the estimation. As for L_c, the classifier was constructed based on the contextual encoder from [15] with fully connected layers 128-32-1. Training was terminated after 10,000 steps with mini-batch size of 4 using the Adam optimizer, which includes 2 labeled and 2 unlabeled patches. The ramp-up weight was applied on τ_1 and τ_2 to control the training process.

Results Analysis: As for the experiments of using DQN as pre-selection, we compared localization accuracy for different volume sizes as environment: 128^3 and 160^3 voxels, since this would use different amounts of contextual informa-

tion within the fixed observation space. The metric is the Euclidean distance between the detected catheter center point and ground-truth center point by voxels in the resolution of 160^3 voxels. Statistical performance of the 128^3 case is 4.3 ± 2.6 voxels, while it becomes 4.7 ± 4.8 voxels for the 160^3 voxels case (measured by mean $\pm$ std.). From the result, 128^3 voxels' environment provides a better detection accuracy without failure because of the increased contextual information that is observed by the agent for a fixed input size.

With the detected catheter center point, patches with size of 48^3 are extracted around the point for semantic segmentation (i.e. 2^3 patches). Segmentation results are obtained based on DQN detection. To evaluate the performance, we consider the Dice Score (DSC), Jaccard Score and 95% Hausdorff Distance (95HD) as the evaluation metrics [11]. We have implemented several state-of-the-art (SOTA) SSL methods for comparison, which include Bayesian UNet [5], Π-model [7] Adversarial-based segmentation (AdSeg)[17], multi-task attention-based SSL (MA-SSL)[2] and uncertainty-aware-based mean-teacher (UA-MT)[16] (all of them are based on Compact-UNet [15]). Results are shown in Fig. 2, which shows the proposed method outperforms SOTA SSL approaches by at least 4% DSC on average. As the number of labeled images increases, the segmentation performances are improved w.r.t available supervised information except for the Π-model, which is due to the unreliable information of the predicted unlabeled images. Because of the unreliable information in challenging 3D US images, the Π-model obtains a lower performance than a simple Bayesian UNet. Compared to MA-SSL and UA-MT, our method achieves a better performance, since it exploits inter/intra-network uncertainty information and boosts the information usage of unlabeled images. To perform ablation studies, the loss components in Eqn. (2) are gradually introduced based on the Dual-UNet (denoted as DU), which are shown in Table. 1. The ablation studies show the proposed constraint component can make use of unlabeled information at different levels, and therefore improve the performances. As can be observed, using 18 labeled images and 42 unlabeled images, the proposed scheme achieves a higher performance than the supervised learning method. From the experiments, the proposed two-stage scheme achieves around 1.2 sec. per volume (0.7+0.5 sec.). As a comparison, exhaustive patch-based segmentation spends > 10 sec. per volume [15], while a voxel-of-interest-based CNN method costs around 10 sec. [14].

4　Conclusion

In this paper, we have proposed a catheter segmentation scheme using a DQN-driven semi-supervised learning for US-guided cardiac intervention therapy. In the proposed method, we design a DQN to coarsely localize the catheter in the 3D US, which is then segmented by an SSL-trained Dual-UNet. With extensive comparison, the proposed method outperforms the SOTAs, while it achieves a higher performance than the supervised learning approach with fewer annotations. Future work will investigate a more complex model with a better supervised loss but to obtain a higher efficiency.

References

1. Alansary, A., Oktay, O., Li, Y., Le Folgoc, L., Hou, B., Vaillant, G., Kamnitsas, K., Vlontzos, A., Glocker, B., Kainz, B., et al.: Evaluating reinforcement learning agents for anatomical landmark detection. Medical image analysis **53**, 156–164 (2019)

2. Chen, S., Bortsova, G., Juárez, A.G.U., van Tulder, G., de Bruijne, M.: Multi-task attention-based semi-supervised learning for medical image segmentation. In: MICCAI. pp. 457–465. Springer (2019)

3. Ghesu, F.C., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., Comaniciu, D.: Multi-scale deep reinforcement learning for real-time 3d-landmark detection in ct scans. IEEE transactions on pattern analysis and machine intelligence **41**(1), 176–189 (2017)

4. Ke, Z., Wang, D., Yan, Q., Ren, J., Lau, R.W.: Dual student: Breaking the limits of the teacher in semi-supervised learning. In: Proceedings of the IEEE CVPR. pp. 6728–6736 (2019)

5. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: NeurIPS. pp. 5574–5584 (2017)

6. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

7. Li, X., Yu, L., Chen, H., Fu, C.W., Heng, P.A.: Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. arXiv preprint arXiv:1808.03887 (2018)

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

9. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: MICCAI. pp. 370–378. Springer (2018)

10. Sedai, S., Antony, B., Rai, R., Jones, K., Ishikawa, H., Schuman, J., Gadi, W., Garnavi, R.: Uncertainty guided semi-supervised segmentation of retinal layers in oct images. In: MICCAI. pp. 282–290. Springer (2019)

11. Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. BMC medical imaging **15**(1), 29 (2015)

12. Yang, H., Shan, C., Kolen, A.F., de With, P.H.N.: Catheter detection in 3d ultrasound using triplanar-based convolutional neural networks. In: 2018 25th IEEE ICIP. pp. 371–375. IEEE (2018)

13. Yang, H., Shan, C., Kolen, A.F., de With, P.H.N.: Automated catheter localization in volumetric ultrasound using 3d patch-wise u-net with focal loss. In: 2019 IEEE ICIP. pp. 1346–1350. IEEE (2019)

14. Yang, H., Shan, C., Kolen, A.F., de With, P.H.N.: Catheter localization in 3d ultrasound using voxel-of-interest-based convnets for cardiac intervention. International journal of computer assisted radiology and surgery **14**(6), 1069–1077 (2019)

15. Yang, H., Shan, C., Tan, T., Kolen, A.F., de With, P.H.N.: Transferring from ex-vivo to in-vivo: Instrument localization in 3d cardiac ultrasound using pyramid-unet with hybrid loss. In: MICCAI. pp. 263–271. Springer (2019)

16. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: MICCAI. pp. 605–613. Springer (2019)

17. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: MICCAI. pp. 408–416. Springer (2017)