SIX-Net: Spatial-context Information miX-up for Electrode Landmark Detection

Xinyi Wang[1,2,3], Zikang Xu[1,2], Heqin Zhu[1,2], Qingsong Yao[4], Yiyong Sun[3], and S. Kevin Zhou[1,2,4]✉

[1] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230026, P.R.China
[2] Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu, 215123, P.R.China
[3] Shanghai MicroPort EP MedTech Co., Ltd. Shanghai, 201318, P.R.China
[4] Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China

Abstract. Catheter ablation is a prevalent procedure for treating atrial fibrillation, primarily utilizing catheters equipped with electrodes to gather electrophysiological signals. However, the localization of catheters in fluoroscopy images presents a challenge for clinicians due to the complexity of the intervention processes. In this paper, we propose SIX-Net, a novel algorithm intending to localize landmarks of electrodes in fluoroscopy images precisely, by mixing up spatial-context information from three aspects: First, we propose a new network architecture specially designed for global-local spatial feature aggregation; Then, we mix up spatial correlations between segmentation and landmark detection, by sequential connections between the two tasks with the help of the Segment Anything Model; Finally, a weighted loss function is carefully designed considering the relative spatial-arrangement information among electrodes in the same image. Experiment results on the test set and two clinical-challenging subsets reveal that our method outperforms several state-of-the-art landmark detection methods ($\sim$ 50% improvement for RF and $\sim$ 25% improvement for CS).

Keywords: Catheter electrode detection · Fluoroscopy analysis

1 Introduction

Atrial Fibrillation (AFib), atrial flutter, and premature ventricular contractions (PVC) are prevalent manifestations of cardiac arrhythmias. Frequent cardiac arrhythmias may give rise to serious consequences, for instance, AFib can lead to blood clots in the heart [19]. Compared with pharmaceutical interventions, catheter-based radiofrequency ablation techniques in cardiac electrophysiology (EP) stand as the standard surgical intervention for the definitive treatment of rapid cardiac arrhythmias, characterized by immediate therapeutic effects and high success rates [11,14]. The electrode constitutes a pivotal component of catheters utilized for both EP signal acquisition and catheter localization.

However, the variability in X-ray imaging quality and the overlap among multiple catheters during clinical surgical procedures render real-time precise electrode localization in fluoroscopy challenging for physicians. To alleviate burdens for clinicians in the surgery and help novice doctors get familiar with this surgery, it is important to develop accurate catheter placement detection methods.

CathSeg [24], FWNet [13] and several works [10,3,2,20,6,17] regard this task as a segmentation problem and require manually annotated masks for supervised training; ConTrack [5] integrates information between video frames and adopts tracking methods. Unlike these methods, in this paper, we model the task as a **single-image electrode landmark detection (SIELD)** problem, where only the input image and ground truth landmark locations are available.

Compared to other SIELD tasks, which focus on landmarks of the discrete structure, such as joints of hands [28] or heads [16,4], the landmarks of the electrodes in intracardiac catheter interventions have a stronger spatial configuration, that is, the electrodes are manufactured to be evenly distributed along the catheter body. How to utilize this **spatial-context information** prior effectively, is the key point for improving the detection precision.

Therefore, in this paper, we propose **SIX-Net** to solve this task by leveraging spatial-context information exhaustively and mixing it up in a deep network for precise SIELD, mainly from three perspectives: **(I) between global and local features:** the global and local features are extracted and aggregated using a U-shape network and a dilated convolution network; **(II) between segmentation masks and landmarks:** we generate pseudo segmentation masks using SAM with ground truth landmarks and leverage the spatial-context information from pseudo masks for landmark detection. The generation of segmentation is low-cost and efficient; **(III) among different landmarks:** by analyzing the spatial distribution of hard-to-detect landmarks in the same image, we witness a significant distribution trend and accordingly modify the weights of the loss function to improve the overall detection precision. This paper offers the following contributions:

1. We propose a landmark detection network, which extracts and integrates the spatial-context features in fluoroscopic images for precise electrode localization;
2. We introduce a weight loss function based on the spatial distribution of hard cases to better tackle clinically challenging situations;
3. Extensive experiments on three datasets illustrate that our proposed method outperforms the state-of-the-art landmark detection methods on two commonly used catheters.

2 Related Works

Catheter Segmentation. FWNet [13] introduces a framework, which combines a segmentor, an optical flow network, and a flow-guided warping function to learn temporal continuity for catheter segmentation in a fluoroscopy sequence.

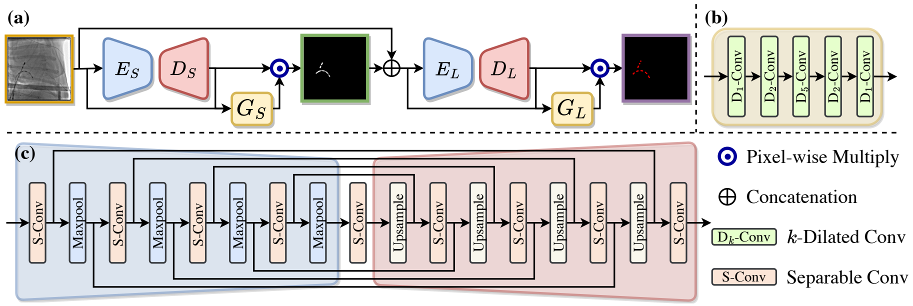

Fig. 1. Overview of SIX-NET. (a) The overall architecture; (b) Architecture of G; (c): Architecture of E and D.

Ambrosini *et al.* [1] emphasize segmentation and center-line construction using both current and preceding images. [24] uses a patch-wise semantic segmentation with model fitting for catheter segmentation in 3D cardiac ultrasound. However, segmenting catheters requires significant effort, and when it comes to subsequent development, the centerline and electrode positions are generally more concerned.

Catheter Tracking. In [21], a tracker is implemented to localize the tip in the last frame as a reference for segmenting the tip in successive frames. U-LanD [7] capitalizes on the uncertainty inherent in landmark prediction to achieve automatic detection of landmarks in key frames of videos. The ConTrack [5] incorporates multiple template images for robustness against appearance changes and employs optical flow computation between frames for refinement.

Landmark Detection. In [25], a multi-task U-Net is implemented to predict both heatmap and offset maps of landmarks simultaneously. In [12], an efficient contour-hugging landmark detection method with uncertainty estimation is depicted. In [28], a universal anatomical landmark detection model has been developed. And OFELIA [23] integrates spatial and temporal features between adjacent frames for electrode localization, aided by optical flow maps.

3 Method

3.1 Problem Definition

Giving a fluoroscopic image $X_i \in \mathbb{R}^{w \times h}$ with the shape of (w, h) and corresponding electrode landmarks $M_i \in \mathbb{R}^{N_e \times w \times h}$, which denotes the position of N_e electrode landmarks, we obtain the k^{th} ($k \in [1, 2, \dots, N_e]$) landmark's heatmap Y_{ik} by using Gaussian function:

$$Y_{ik}(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x - x_{ik})^2 + (y - y_{ik})^2}{2\sigma^2}). \quad (1)$$

Electrode landmark detection aims to train a network $f(\cdot)$, which takes the X_i as input, and predicts the locations of electrode landmarks in it, i.e., $\{\hat{Y_{ik}}\}_{k=1}^{N_e}$.

3.2 Global ↔ Local Mix-up

The global information and local information are extracted by a U-shape net and a dilated convolution net [22]. Specifically, we use E and D to represent the encoder and decoder of the U-Net to extract the local information.

Inspired by [15,9,27], E consists of four separable convolution blocks, termed S-Conv, which is mainly composed of depth-wise convolution and point-wise convolution layers. Each S-Conv is followed by a max-pooling layer to reduce the shape of the feature map. Similarly, D is composed of four up-sampling + S-Conv blocks, trying to predict the pseudo mask or target landmark heatmap. We use the S-Conv because it's lightweight. Skip connections are added between S-Convs of E and E within the same feature levels. The detailed architectures of E and D are shown in Fig. 1.

The global feature is extracted by dilated convolution net G. As shown in Fig. 1, G is a 5-layer convolution net with a dilated ratio of 1, 2, 5, 2, and 1, respectively. The local and global features are aggregated by pixel-wise multiplication before being sent to the next module. For simplification, we use ϕ to represent the whole global-local network. ϕ_S and ϕ_L denote the networks used for segmentation and SIELD, which are of the same architecture.

3.3 Segmentation ↔ Detection Mix-up

Witnessing the fact that the electrode landmarks are concentrated along the catheter, we try to aggregate the segmentation of the catheter to the SIELD task. However, it is hard to access the ground truth segmentation mask due to the heavy annotation workload. Thanks to the development of Segmentation Anything Models [8] (SAM), zero-shot image segmentation with minimum prompt information such as points or bounding boxes would be possible.

In the training stage, we first generate a pseudo segmentation mask S_i using SAM by using the original image X_i and the ground truth landmarks Y_{ik}. i.e. $S_i = \text{SAM}(X_i, Y_{ik})$. The pseudo mask need not be very precise as we only require an approximate estimation of the catheter's shape. Then, we train a segmentor ϕ_S using the aforementioned architecture to predict $\hat{S}_i$. After that, the estimated $\hat{S}_i$ is concatenated to the input X_i, and passed to ϕ_L to predict the final heatmap.

By introducing extra shape features to the landmark localization network, ϕ_L can aggregate the spatial-context information between the catheter and electrodes and improve the overall performance.

3.4 Inter-Electrode Mix-up

In the previous studies, the landmarks are treated equally, i.e. the overall cross-entropy loss of each landmark is computed and averaged with the same importance, as shown below:

$$\mathcal{L}_D = \sum_{k=1}^{N_e} \frac{1}{N_e} \cdot \mathcal{L}_{CE}(Y_k, \hat{Y}_k). \tag{2}$$

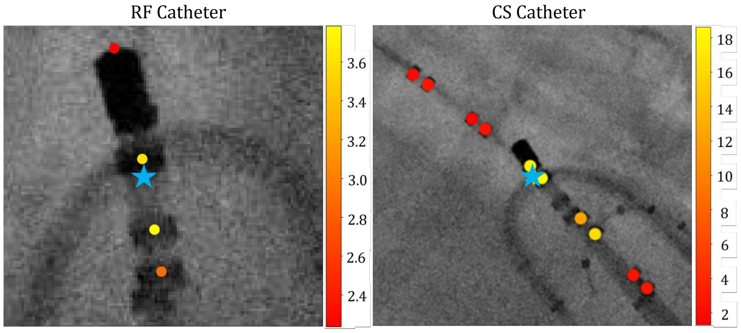

Fig. 2. MRE for each landmark of a sample image. The nearer electrode landmarks have higher prediction errors. ●●●: ground truth landmarks with different MRE; ⋆: central point of landmarks. Left: RF Catheter; Right: CS Catheter.

However, the hardness of accurate landmark detection varies significantly among the N_e electrodes. Let $Y_{cc} \leftarrow (x_{cc}, y_{cc})$ represent the central coordinate of the N_e electrode landmarks. For current landmark $Y_k \rightarrow (x_k, y_k)$, the closer Y_k to Y_{cc}, the larger prediction error it has, as shown in Fig. 2.

Considering this special spatial-context information, we use different weight factors for each landmark, making ϕ_L pay more attention to the hard landmarks. Specifically, we compute the Euclidean distance between Y_k and Y_{cc} as Equ. 3:

$$\mathcal{D}_{k \leftrightarrow cc} = \sqrt{(x_k - x_{cc})^2 + (y_k - y_{cc})^2}. \tag{3}$$

Then, the weighting factor ω_k is computed using Equ. 4, and the overall loss function $\mathcal{L}_D$ is defined as the weighted sum.

$$\mathcal{L}_D^* = \sum_{k=1}^{N_e} \omega_k \cdot \mathcal{L}_{CE}(Y_k, \hat{Y}_k), \quad \omega_k = \frac{\sqrt{\sum_{i=1}^{N_e} (D_{i \leftrightarrow cc})^2} - D_{k \leftrightarrow cc}}{\sqrt{\sum_{i=1}^{N_e} (D_{i \leftrightarrow cc})^2}}. \tag{4}$$

Note that in our task, each image X_i consists of two types of catheters: radio-frequency (RF) and coronary sinus (CS). Thus, each landmark Y_k is first categorized into one class based on the relative relation between $D_{k \leftrightarrow cc[RF]}$ and $D_{k \leftrightarrow cc[CS]}$, where $Y_{cc[RF]}$ and $Y_{cc[CS]}$ are the central point of RF catheter and CS catheter. Then, the L_D of both catheters are computed and added together.

Compared to the original loss function $\mathcal{L}_D$, this new spatial-context information-aware loss function $\mathcal{L}_D^*$ can provide better differentiation of the boundaries and spatial-context information among electrodes and serve as a strong prior knowledge for landmark detection. Finally, the overall loss function considering the segmentation accuracy and detection precision are formulated as Equ. 5:

$$\mathcal{L} = \mathcal{L}_{\mathcal{S}} + \mathcal{L}_D^* = \mathcal{L}_{CE}(\hat{S}_i, S_i) + \sum_{k=0}^{K} w_{ik} \mathcal{L}_k. \tag{5}$$

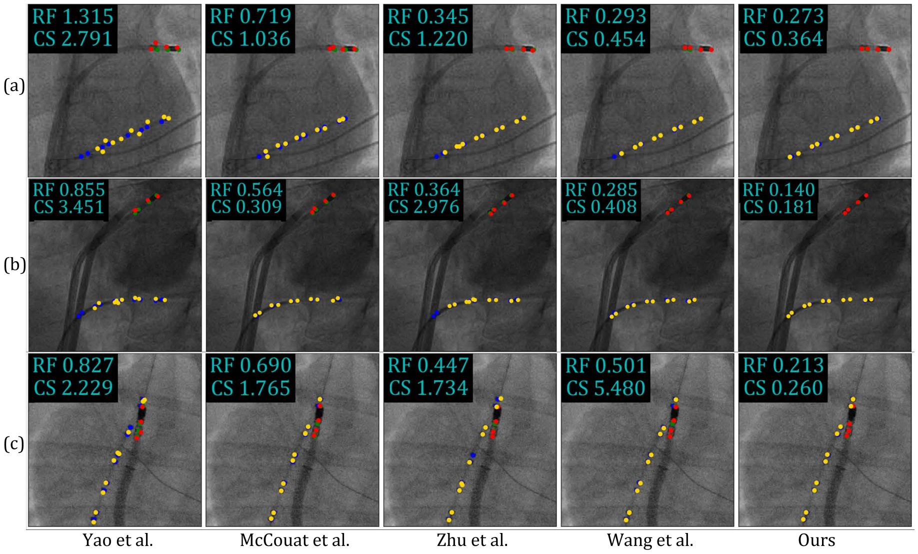

Fig. 3. Qualitative results on the three test sets. The numbers show the MRE of RF and CS catheters. The ground truth and predicted landmark of CS Catheter are in blue and yellow. The ground truth and predicted landmark of RF Catheter are in green and red.

4 Experiments and Results

Experiment Settings about datasets, metrics, and implementation details.
Datasets. This study uses a private multi-center dataset of fluoroscopic sequences obtained from cardiac ablation procedures and animal experiments. The dataset includes two commonly used catheters: CS and RF catheters. Electrode landmarks within this study are identified based on the electrode's center point, except for the RF catheter's initial landmark, which is defined as its tip, resulting in a total of 14 landmarks per frame (4 for RF and 10 for CS). Annotation of the dataset was conducted by two experienced engineers employing the LabelMe tool [18] and reviewed by three clinical experts. The training and test sets consist of 14,768 and 7,711 frames, respectively. To evaluate the robustness and adaptability of the proposed methodology, we extract two clinical-challenging (CCA) subsets, which consist of frames of specific procedural scenes, the injection of contrast agent(575 frames, termed as DSA-Test) and frames where catheters are partially obscured (2,266 frames, denoted as OBS-Test), both of which present augmented complexity for the detection of catheter electrodes.

Metrics We use mean radial error (MRE) to measure the Euclidean distance between prediction and ground truth. Additionally, the Successful Detection Rate (SDR) is determined at four distinct thresholds: 0.5 mm, 1 mm, 2 mm, and 4 mm, to assess the detection precision within these specified radii.

Table 1. Results on the three test sets. **Best** and Second Best are highlighted.

Model	RF Catheter					CS Catheter				
	MRE ↓	SDR (% ↑)				MRE ↓	SDR (% ↑)			
	(mm)	0.5mm	1mm	2mm	4mm	(mm)	0.5mm	1mm	2mm	4mm
Test Dataset										
Yao *et al.* [26]	3.6232	14.19	42.18	48.54	70.03	4.7875	17.12	39.22	65.55	84.91
McCouat *et al.* [12]	0.9955	35.51	90.68	96.02	96.41	0.9237	37.17	92.13	96.59	97.25
Zhu *et al.* [28]	1.2992	78.46	93.78	96.17	96.81	0.9045	**86.80**	92.04	93.00	94.51
OFELIA [23]	0.9175	76.24	94.29	97.71	98.18	0.6628	73.43	96.89	98.60	99.30
SIX-Net(Ours)	**0.4441**	**83.86**	**98.23**	**99.31**	**99.57**	**0.4327**	84.99	**98.96**	**99.56**	**99.80**
Test-DSA Dataset										
Yao *et al.* [26]	5.7585	18.69	38.16	76.55	79.52	3.5116	27.60	65.51	67.97	75.47
McCouat *et al.* [12]	1.7882	40.35	83.65	91.48	94.96	0.6143	44.35	93.04	98.96	99.83
Zhu *et al.* [28]	4.6221	43.48	83.83	91.83	94.26	0.5424	**85.74**	94.09	96.17	97.22
OFELIA [23]	1.6765	60.52	89.57	92.52	95.48	0.6127	64.33	94.26	99.65	**100.00**
SIX-Net(Ours)	**0.5713**	**68.00**	**94.78**	**97.04**	**98.96**	**0.4056**	83.13	**97.91**	**99.83**	**100.00**
Test-OBS Dataset										
Yao *et al.* [26]	4.5076	35.17	47.33	58.53	83.08	3.0357	25.99	39.98	43.14	80.01
McCouat *et al.* [12]	1.0378	43.93	93.60	96.95	97.57	0.8304	42.56	93.42	97.35	98.41
Zhu *et al.* [28]	1.1779	81.94	95.41	96.95	97.88	0.9970	84.15	88.92	89.89	93.02
OFELIA [23]	0.9586	76.36	93.91	97.00	98.01	0.6493	64.59	95.32	97.62	99.16
SIX-Net(Ours)	**0.4739**	**83.49**	**97.44**	**98.32**	**99.16**	**0.4854**	**86.58**	**97.35**	**99.26**	**99.66**

Implementation details. Our model is implemented in PyTorch and trained on an NVIDIA A100 GPU. The image pairs are augmented by random rotation, intensity scaling, and elastically transformation, and resized to 640 × 640 before being sent to the network. The network training is conducted utilizing the Adam optimizer, commencing with a learning rate of 0.001 and employing a batch size of 4 for 40 epochs. Learning rate adjustments are implemented by decreasing it by a factor of 0.1 at epochs 4, 8, 12, 16, and 32.

Results We compare SIX-NET with several commonly used algorithms for medical landmark detection [25,12,28,23], and the quantitative results are shown in Table 1. It is observed that our SIX-NET outperforms the four SOTA on almost all metrics on the test sets. This demonstrates the value of spatial-context information mixup from multiple perspectives, as the other methods mainly focus on the spatial feature learned from a single level. Besides, our method presents good generalization on the two CCA test sets (The SDR drop is much lower compared to other methods), which is a justifiable phenomenon as bringing in extra knowledge improves the robustness of the network and provides an aid to deal with difficult situations. We also present qualitative results of different detection methods in Fig. 3, and our method outperforms other methods significantly.

Ablation Study Ablation studies are conducted to evaluate the usefulness of different modules of SIX-Net, i.e. Global-Local Mixup (GL-Mixup), Segmentation-

Table 2. Sub-Module ablation. **Best** and Second Best are highlighted.

Module			RF Catheter					CS Catheter				
GL	SD	IE	MRE ↓	SDR (% ↑)				MRE ↓	SDR (% ↑)			
Mixup	Mixup	Mixup	(mm)	0.5mm	1mm	2mm	4mm	(mm)	0.5mm	1mm	2mm	4mm
	✓	✓	0.5839	62.50	95.56	98.85	99.39	0.6369	71.96	95.26	97.83	98.41
✓		✓	0.8358	63.18	94.28	97.21	98.08	0.9808	77.05	95.14	97.84	98.51
✓	✓		0.9448	62.66	94.73	97.83	98.22	0.6523	78.57	96.11	98.13	98.67
✓	✓	✓	**0.4441**	**83.86**	**98.23**	**99.31**	**99.57**	**0.4327**	**84.99**	**98.96**	**99.56**	**99.80**

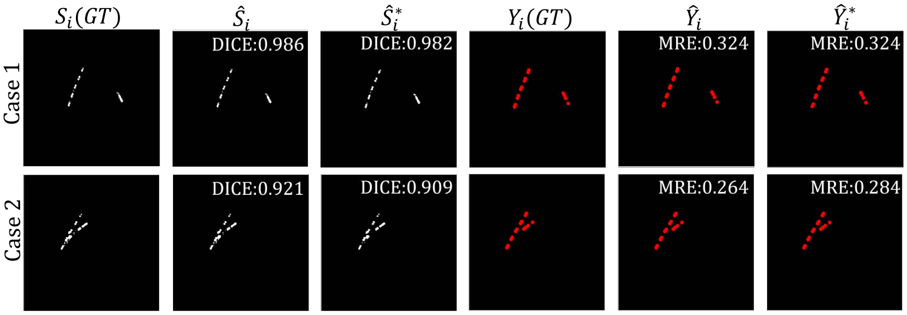

Fig. 4. The results of bi-direction verification. S_i: pseudo ground truth mask generated by SAM. $\hat{S}_i$: output of $\phi_S(X_i)$). $\hat{S}_i^*$: output of SAM using X_i and predicted landmarks $\hat{Y}_i$. Y_i: ground truth landmarks. $\hat{Y}_i$: the predicted landmarks. $\hat{Y}_i^*$: output of SIX-Net when replacing $\hat{S}_i$ with S_i.

Detection Mixup (SD-Mixup), and Inter-Electrode Mixup (IE-Mixup). The following changes are modified on SIX-Net: (I) Replacing GL network with origin U-Net; (II) Only use ϕ_D; (III) Using L_D instead of L_D^*. The results are shown in Table 2. Besides, we also conduct bi-direction verification on SIX-Net. **(I) Backward direction:** We use the predicted landmarks $\hat{Y}_i$ as the prompt for SAM and compare the segmentation $\hat{S}_i^* = \mathrm{SAM}(X_i, \hat{Y}_i)$ with S_i; **(II) Forward Direction:** We use S_i to replace $\hat{S}_i$ and compare the predicted landmark $\hat{Y}_i^*$ with Y_i. The results are shown in Figure 4. The similar Dice between $\hat{S}_i$ and $\hat{S}_i^*$ illustrates the backward precision of SIX-Net and the similar MRE between $\hat{Y}_i$ and $\hat{Y}_i^*$ proves the forward precision of SIX-Net.

5 Conclusion and Future Work

Accurate and efficient electrode detection in real-time fluoroscopy holds paramount significance. In this work, we propose SIX-Net, which mixes up three-level spatial-context information: between global and local features, between segmentation maps and landmark heatmaps, and inter-electrode arragement, into the pipeline for precise electrode localization in X-ray images. The results on the test set and two CCA subsets illustrate the efficiency of our proposed SIX-NET compared with several SOTA methods. Further research could be conducted on the exploration of one-shot or few-shot methods to alleviate the burden of electrode annotation and discover some efficient and low-cost methods for other types of catheters.

Acknowledgments. This study was funded by Natural Science Foundation of China under Grant 62271465 and Open Fund Project of Guangdong Academy of Medical Sciences, China (No.YKYKF202206).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ambrosini, P., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: Fully automatic and real-time catheter segmentation in x-ray fluoroscopy (2017)
2. Bodart, L.E., Ciske, B.R., Le, J., Reilly, N.M., Deaño, R.C., Ewer, S.M., Tipnis, P., Rahko, P.S., Wagner, M.G., Raval, A.N., Speidel, M.A.: Technical and clinical study of x-ray-based surface echo probe tracking using an attached fiducial apparatus. Medical Physics (2020)
3. Chang, P.L., Rolls, A., Praetere, H.D., Poorten, E.V., Riga, C.V., Bicknell, C.D., , Stoyanov, D.: Robust catheter and guidewire tracking using b-spline tube model and pixel-wise posteriors. IEEE Robotics and Automation Letters (2016)
4. Chen, R., Ma, Y., Liu, L., Chen, N., Cui, Z., Wei, G., Wang, W.: Semi-supervised anatomical landmark detection via shape-regulated self-training. Neurocomputing **471**, 335–345 (2022)
5. Demoustier, M., Zhang, Y., Narasimha Murthy, V., Ghesu, F.C., Comaniciu, D.: Contrack: Contextual transformer for device tracking in x-ray. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 679–688 (2023)
6. Huang, L., Liu, Y., Chen, L., Chen, E.Z., Chen, X., Sun, S.: Robust landmark-based stent tracking in x-ray fluoroscopy. In: Computer Vision – ECCV 2022. pp. 201–216. Cham (2022)
7. Jafari, M.H., Luong, C., Tsang, M., Gu, A.N., Van Woudenberg, N., Rohling, R., Tsang, T., Abolmaesumi, P.: U-land: Uncertainty-driven video landmark detection. IEEE Transactions on Medical Imaging **41**(4), 793–804 (2022). https://doi.org/10.1109/TMI.2021.3123547
8. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything (2023)
9. Lian, C., Wang, F., Deng, H.H., Wang, L., Xiao, D., Kuang, T., Lin, H.Y., Gateno, J., Shen, S.G., Yap, P.T., et al.: Multi-task dynamic transformer network for concurrent bone segmentation and large-scale landmark localization with dental cbct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 807–816. Springer (2020)
10. Ma, Y., Gogin, N., Cathier, P., Housden, R.J., Gijsbers, G., Cooklin, M., O'Neill, M., Gill, J., Rinaldi, C.A., Razavi, R., Rhode, K.S.: Real-time x-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions. Medical Physics (2013). https://doi.org/10.1118/1.4808114

11. Mark, D.B., Anstrom, K.J., Sheng, S., Piccini, J.P., Baloch, K.N., Monahan, K.H., Daniels, M.R., Bahnson, T.D., Poole, J.E., Rosenberg, Y., Lee, K.L., Packer, D.L., for the CABANA Investigators: Effect of Catheter Ablation vs Medical Therapy on Quality of Life Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. JAMA **321**(13), 1275–1285 (04 2019). https://doi.org/10.1001/jama.2019.0692, https://doi.org/10.1001/jama.2019.0692
12. McCouat, J., Voiculescu, I.: Contour-hugging heatmaps for landmark detection. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20565–20573 (2022). https://doi.org/10.1109/CVPR52688.2022.01994
13. Nguyen, A., Kundrat, D., Dagnino, G., Chi, W., Abdelaziz, M.E.M.K., Guo, Y., Ma, Y., Kwok, T.M.Y., Riga, C., Yang, G.Z.: End-to-end real-time catheter segmentation with optical flow-guided warping during endovascular intervention. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 9967–9973 (2020). https://doi.org/10.1109/ICRA40945.2020.9197307
14. Parameswaran, R., Al-Kaisey, A.M., Kalman, J.M.: Catheter ablation for atrial fibrillation: current indications and evolving technologies. Nature Reviews Cardiology **18**(3), 210–225 (2021)
15. Payer, C., Štern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based cnns for landmark localization. Medical image analysis **54**, 207–219 (2019)
16. Quan, Q., Yao, Q., Li, J., Zhou, S.K.: Which images to label for few-shot medical landmark detection? (2021)
17. Ramadani, A., Bui, M., Wendler, T., Schunkert, H., Ewert, P., Navab, N.: A survey of catheter tracking concepts and methodologies. Medical Image Analysis **82**, 102584 (2022). https://doi.org/https://doi.org/10.1016/j.media.2022.102584, https://www.sciencedirect.com/science/article/pii/S1361841522002225
18. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International journal of computer vision **77**, 157–173 (2008)
19. Staerk, L., Sherer, J.A., Ko, D., Benjamin, E.J., Helm, R.H.: Atrial fibrillation: Epidemiology, pathophysiology, and clinical outcomes. Circulation Research **120(9)**, 1501–1517 (2017)
20. Torabinia, M., Caprio, A., Jang, S.J., Ma, T., Tran, H., Mekki, L., Chen, I., Sabuncu, M.R., Wong, S.C., Mosadegh, B.: Deep learning-driven catheter tracking from bi-plane x-ray fluoroscopy of 3d printed heart phantoms. Mini-invasive Surgery (2021), https://api.semanticscholar.org/CorpusID:237815143
21. Ullah, I., Chikontwe, P., Park, S.H.: Real-time tracking of guidewire robot tips using deep convolutional neural networks on successive localized frames. IEEE Access **7**, 159743–159753 (2019). https://doi.org/10.1109/ACCESS.2019.2950263
22. Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.: Understanding convolution for semantic segmentation. In: 2018 IEEE winter conference on applications of computer vision (WACV). pp. 1451–1460. IEEE (2018)
23. Wang, X., Xu, Z., Yao, Q., Sun, Y., Zhou, S.K.: OFELIA: Optical flow-based electrode localization. In: Submitted to Medical Imaging with Deep Learning (2024), https://openreview.net/forum?id=8245ExLB4I, under review
24. Yang, H., Shan, C., Kolen, A.F., N. de With, P.H.: Automated catheter localization in volumetric ultrasound using 3d patch-wise u-net with focal loss. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 1346–1350 (2019). https://doi.org/10.1109/ICIP.2019.8803045

25. Yao, Q., He, Z., Han, H., Zhou, S.K.: Miss the point: Targeted adversarial attack on multiple landmark detection (2020)
26. Yao, Q., Quan, Q., Xiao, L., Zhou, S.K.: One-shot medical landmark detection (2021)
27. Zhu, H., Yao, Q., Xiao, L., Zhou, S.K.: You only learn once: Universal anatomical landmark detection. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 85–95. Springer (2021)
28. Zhu, H., Yao, Q., Xiao, L., Zhou, S.K.: Learning to localize cross-anatomy landmarks in x-ray images with a universal model. BME frontiers **2022** (2022)