Review of the paper : *Using Attention-based Convolutional Auto-Encoders for Catheter Path Reconstruction in Ultrasound Images*

Raphaël Faure and Nassim Arifette

March 14, 2025

1 Paper Summary

Prostate cancer is the second most prevalent cancer among men and the fifth leading cause of mortality globally[8]. One way of treatment is High-dose-rate (HDR) brachytherapy, which often involves the insertion of catheters under image guidance. Knowing the precise location of catheters is critical to ensure accurate radiation dose delivery and minimize damage to surrounding healthy tissue. The manual procedure of catheter identifications compels clinicians to scroll through 3D TRUS (Transcretal ultrasound) images. However, these images are low signal-to-noise making identifications difficult and time-consuming. Deep-learning methods have been recently developed as automated approaches for catheter path reconstruction and this article [7] goes in this sens with presenting an innovative method.

Deep learning methods represent almost the state-of-the-art. Some solutions based on electromagnetic methods were proposed but they were limited due to electromagnetic interferences and noises. [6] used deep learning and the 3D Hough transform for catheter localization, achieving high accuracy but needing improvements in tip localization. [5] applied feature point classification and trajectory refinement on 3D ultrasound images for catheter segmentation. More recently, [1] and [9] enhanced a U-Net architecture with attention gates and total variation regularization, improving prediction accuracy and reducing localization errors but still struggling with a noisy environment. The authors propose a new method based on attention-based convolutional auto-encoders trained on an augmented dataset from which they obtain a better global accuracy, good robustness to noise and complex trajectories, and better inference. The last point highlights how this innovative method could be used in real-time identifications.

The proposed solution is based on the following pipeline: it takes 3D ultrasound images as input and processes them through an attention-based autoencoder to generate 2D binary segmentation masks of catheter locations. These segmentations are refined using contour detection and centroid calculation to form a 3D point cloud, which is then used by a modified RANSAC algorithm to reconstruct the 3D catheter paths.

This approach is mostly based on two valued points : data augmentation and the model's architecture. Building datasets are always complicated when it deals with medical images. On top of that, previous methods were struggling with complex and noisy cases. In order to fix these issues, the authors use Faster AutoAugment [4] to perform an optimized data augmentation on their dataset which increased their number of images to 5830 compared to 1480 before. The training and validation datasets were augmented using three techniques: affine transformations to account for varying catheter orientations, perspective transformations to simulate different imaging viewpoints, and pixel-level transformations to improve generalization under challenging imaging conditions. On the other hand, the proposed architecture offers two significant advantages. Firstly, it combines an autoencoder with spatial and channel attention mechanisms, allowing it to focus on the most relevant features in ultrasound images. This not only improves the accuracy of catheter segmentation but also provides better interpretability by highlighting important regions of the image. Secondly, this architecture is designed to be lightweight and fast, making it an ideal solution for real-time use in clinical settings.

To train their model, they use free TRUS 3D images on which catheters were synthetically inserted [3][2], then they perform data augmentation. Once they have this new dataset, they split it into 3 parts: training set, validation set, and test set as usual. The model was trained from scratch on a single NVIDIA A100 GPU for 100 epochs, using Kaiming Initialization and the Adam optimizer with weight decay. The Focal Tversky loss function was employed, with optimized parameters α, β, and γ. The learning rate started at 0.001 and decreased after 50 epochs. The initial convolutional layer had 16 channels for efficiency, and a drop-block discard size of 7 was used with a linear drop rate schedule. The performance was evaluated against a U-Net model using metrics including catheter shaft and tip localization error, detection accuracy, precision, recall, Dice score, IoU, and MCC.

This solution achieved a high catheter path detection rate of approximately 98 %, with mean tip and shaft errors of 0.18 mm and 0.39 mm, respectively, depending on catheter path complexity. Compared to the U-Net architecture, the proposed model demonstrated superior performance, achieving a higher detection accuracy (97.954% vs. 93.143%) and lower mean tip (0.178 mm) and shaft (0.384 mm) localization errors. Furthermore, the proposed model outperformed U-Net in Dice score, precision, IoU, and MCC, with a notably faster average inference time of 0.0029 seconds.

2 Critical assessment

The authors successfully combine attention mechanisms, auto-augmentation, and a modified RANSAC algorithm in a novel manner that specifically targets the challenges of catheter detection in noisy ultrasound images. The attention mechanism implementation is well suited to address the low signal-to-noise ratio inherent in ultrasound imaging.

The computational efficiency shown is quite impressive and clinically relevant. The reported inference speed ($0.0029s$) is approximately twice as fast as the U-Net baseline ($0.0057s$) on comparable hardware. This lightweight architecture with just 16 channels in the initial convolutional layer is well designed for real-time clinical requirements. On top of that, the quantitative results are strong with 98% detection accuracy and low mean errors (tip: 0.18 ± 0.12mm, shaft: 0.39 ± 0.28mm). The comprehensive evaluation using multiple metrics (Dice, precision, recall, IoU, MCC) provides a thorough assessment of performance. The visual results effectively illustrate the model's capabilities.

The most significant limitation, acknowledged by the authors, is the reliance on an extremely limited dataset of only three base 3D TRUS images, from which 51 modified images were generated with synthetically inserted catheters. It raises questions about generalizability to real clinical scenarios. The paper states: *"These images were then modified to simulate the presence of implanted catheters."* However, synthetic simulations may not fully represent the artifacts, noise patterns, and complex interactions that occur with real catheters in patient tissues. Validation on a dataset containing real clinical images with actual inserted catheters from multiple patients and institutions is essential to demonstrate clinical viability.

The comparison with existing methods requires enhancement. While the paper cites several recent approaches in the introduction [6, 1, 9, 10], it only compares the proposed method against a standard U-Net architecture. This is insufficient to establish advancement over current state-of-the-art approaches. For instance, the authors note that [6] achieved 95% accuracy and [1, 9] improved prediction accuracy to 96%, but do not provide direct comparisons of tip and shaft errors.

The paper lacks ablation studies to demonstrate the contribution of each component. Without this analysis, it is difficult to assess which aspects (attention mechanism, auto-augmentation technique, modified RANSAC algorithm) are most innovative and impactful.

The paper reports that performance varies *"depending on the complexity of the catheter's curve path"* but doesn't provide a systematic analysis of what factors contribute to higher errors. There's no detailed investigation of failure cases or challenging scenarios, which would be valuable for understanding the method's limitations and reliability. A more thorough error analysis examining factors such as catheter depth, angle, and proximity to anatomical structures would provide valuable insights into the method's limitations.

The authors provide specific hyperparameter values ($\alpha = 0.7$, $\beta = 0.3$, $\gamma = 1.33$ for Focal Tversky loss, dropout probability of 0.2, etc.) without explanation of how these values were determined. The absence of hyperparameter optimization details or sensitivity analysis makes reproducibility challenging and obscures understanding of the method's robustness.

While the paper mentions potential to *"improve the prostate brachytherapy workflow"*, it doesn't adequately discuss practical aspects of integration into existing clinical systems and protocols. A small user study involving radiation oncologists would strengthen the clinical relevance. Additionally, there is no discussion of how variations in ultrasound machine settings or image acquisition parameters might affect the algorithm's performance, nor how different catheter materials or designs might impact detection accuracy.

The inference is reported for an NVIDIA A100 GPU, which may not be realistic for clinically accessible hardware. Evaluating performance on computing resources more commonly available in clinical settings would provide more realistic expectations for implementation.

To strengthen this work, we recommend conducting validation studies with real patient data containing actual implanted catheters, performing comprehensive comparisons with recent state-of-the-art methods, and including ablation studies to isolate the contributions of individual components. Additionally, the authors should analyze failure cases systematically, evaluate performance on standard clinical hardware, and discuss integration into clinical workflows more thoroughly. A clearer explanation of hyperparameter selection and testing the robustness of the method under varying imaging conditions would also significantly enhance the paper's contribution.

References

[1] X. Dai et al. "Automatic multi-catheter detection using deeply supervised convolutional neural network in mri-guided hdr prostate brachytherapy". In: *Medical Physics* 47.9 (2020), pp. 4115–4124.

[2] A. Fedorov et al. *Annotated MRI and ultrasound volume images of the prostate*. Mar. 2015. DOI: 10.5281/zenodo.16396. URL: https://doi.org/10.5281/zenodo.16396.

[3] A. Fedorov et al. "Open-source image registration for mri–trus fusion-guided prostate interventions". In: *International journal of computer assisted radiology and surgery* 10 (2015), pp. 925–934.

[4] R. Hataya et al. "Faster autoaugment: Learning augmentation strategies using backpropagation". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV*. Vol. 16. Springer, 2020, pp. 1–16.

[5] W. T. Hrinivich et al. "Simultaneous automatic segmentation of multiple needles using 3d ultrasound for high-dose-rate prostate brachytherapy". In: *Medical physics* 44.4 (2017), pp. 1234–1245.

[6] N. Kitner et al. "Multicatheter modelling in reconstructed 3d transrectal ultrasound images from prostate brachytherapy". In: *Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 12466. SPIE. 2023, pp. 126–135.

[7] Shreyasi Mandal, Srinjoy Bhuiya, and Elodie Lugez. "Using attention-based convolutional auto-encoders for catheter path reconstruction in ultrasound images". In: *Medical Imaging 2024: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Jeffrey H. Siewerdsen and Maryam E. Rettmann. Vol. 12928. International Society for Optics and Photonics. SPIE, 2024, 129281N. DOI: 10.1117/12.3003897. URL: https://doi.org/10.1117/12.3003897.

[8] P. Rawla. "Epidemiology of prostate cancer". In: *World journal of oncology* 10.2 (2019), p. 63.

[9] Y. Zhang et al. "Multi-needle localization with attention u-net in us-guided hdr prostate brachytherapy". In: *Medical physics* 47.7 (2020), pp. 2735–2745.

[10] Y. Zhao, C. Cachard, and H. Liebgott. "Automatic Needle Detection and Tracking in 3D Ultrasound Using an ROI-Based RANSAC and Kalman Method". In: *Ultrasonic Imaging* 35.4 (2013), pp. 283–306.