Using Attention-based Convolutional Auto-Encoders for Catheter Path Reconstruction in Ultrasound Images

Shreyasi Mandal[1*], Srinjoy Bhuiya[2*], and Elodie Lugez[3]

[1]Indian Institute of Technology, Kanpur, India
[2]University of Alberta, Edmonton, Canada
[3]Toronto Metropolitan University, Toronto, Canada
[*]These authors contributed equally to this work.

ABSTRACT

Purpose: In high-dose-rate prostate brachytherapy, multiple catheters are inserted into the prostate gland. Conventional catheter detection in ultrasound images is hindered by images of a low signal-to-noise ratio, making it difficult to identify the catheters manually. To address this issue, this paper presents an innovative automated deep-learning approach for catheter path reconstruction.

Method: A lightweight spatial attention-based autoencoder convolutional neural network is developed to accurately and rapidly segment catheters in real-time volumetric transrectal ultrasound images. To overcome the challenges of noisy and limited annotated data, an auto-augmentation technique is employed, which leverages a controller network to learn the optimal augmentation policy. A 3D random sample consensus-based catheter path reconstruction algorithm is also proposed to transform the catheter segmentations into smooth curves representing their paths.

Results: By integrating automated data augmentation with an optimized autoencoder network, structured dropout, and batch normalization techniques, the proposed algorithm successfully detected 98% of the catheter paths tested. The mean tip errors were recorded as 0.18 ± 0.12 mm, while the mean shaft errors were recorded as 0.39 ± 0.28 mm, varying depending on the complexity of the catheter curve path. Notably, the proposed algorithm outperformed existing methods by exhibiting faster inference times, with an average inference time of 0.0029 seconds.

Conclusion: The proposed methodology offers a comprehensive approach to enhance the accuracy of catheter path reconstruction in prostate brachytherapy. This lightweight neural network also has the potential to significantly improve the prostate brachytherapy workflow by making the catheter reconstruction process time-efficient.

Keywords: Deep Learning, Ultrasound, Segmentation, Convolutional Neural Network, High-Dose-Rate Brachytherapy

1. INTRODUCTION

Prostate cancer is the second most common cancer affecting the male population and the fifth leading cause of death worldwide.[1] High-dose-rate (HDR) brachytherapy has emerged as a pivotal treatment option for prostate cancer since the 1980s.[2] It usually involves the insertion of catheters under image guidance.[3] Accurate reconstruction of the three-dimensional (3D) trajectory of these catheters, with respect to the prostate, is crucial to ensure precise dose distribution while minimizing damage to healthy tissues.[4] The existing manual procedure for catheter identification in HDR prostate brachytherapy relies on clinicians manually scrolling through three-dimensional transrectal ultrasound (TRUS) images, which is often tedious and time-consuming.
Recent integrations of imaging modalities, real-time tracking systems, and advanced algorithms have paved the way for accurate and efficient catheter path reconstructions. Indeed, electromagnetic (EM) tracking has been

Further author information: Elodie Lugez
E-mail: elodie.lugez@torontomu.ca

Medical Imaging 2024: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Jeffrey H. Siewerdsen, Maryam E. Rettmann, Proc. of SPIE Vol. 12928,
129281N · © 2024 SPIE · 1605-7422 · doi: 10.1117/12.3003897

proposed for catheter localization. However, it is susceptible to electromagnetic interference and noise.[5] Although efforts have been made to improve EM measurements with the development of filters,[6,7] their clinical implementation is still pending. Kitner et al.[8] proposed a two-step pipeline combining deep learning and the 3D Hough transform to accurately locate catheters in HDR brachytherapy. Although the pipeline achieved a high accuracy of 95%, with an average difference of 1.9 mm for catheter shafts and 3.0 mm for catheter tips compared to ground truth positions, further improvements were needed to enhance tip localization and handle complex catheter paths. Hrinvich et al.[9] applied regularized feature point classification, trajectory identification and refinement on reconstructed 3D ultrasound images to segment and identify catheters' trajectories. Dai et al.[10] and Zhang et al.[11] incorporated attention gates and total variation regularization to a U-Net architecture, improving prediction accuracy to 96% with small shaft and tip localization errors. Zhao et al.[12] used a line filter to enhance catheter contrast. A random sample consensus-based (RANSAC) algorithm and Kalman filters were then applied to accurately detect and track the catheter position. However, the proposed method failed in noisy backgrounds and cases of significant relative motion between the probe and catheter.

This paper presents a lightweight Convolutional Neural Network (CNN) for real-time catheter path reconstruction in TRUS images. The proposed system relies on an attention-based convolutional autoencoder incorporating channel and spatial attention mechanisms. These enable enhanced segmentation accuracy by focusing on informative features and reducing the impact of noise and artifacts in the images. Furthermore, an auto-augmentation technique was employed to improve generalization and reduce overfitting. Finally, the efficacy of our proposed methods will be compared against another state-of-the-art model using the same dataset.

2. MATERIALS AND METHODS

2.1 Dataset and Data Augmentation

This study relies on real TRUS images of the prostate on which HDR catheters were synthetically inserted. Indeed, the resulting images were generated based on three 3D TRUS images freely available online with no catheters implanted.[13,14] These images were then modified to simulate the presence of implanted catheters. Random simulations of 12 to 25 catheter paths were generated on these images, resulting in a set of 51 3D images. Each 3D TRUS image comprises a stack of thirty-two 2D grayscale images, each with a resolution of 512 x 512 pixels. These images were further transformed into their corresponding Cartesian plane representations and used for training the model.

Faster AutoAugment[15] is applied to our dataset, which learns an optimized augmentation policy by integrating its search pipeline into the model's training process. Three data augmentation techniques are utilized: affine transformations to learn catheter appearance from multiple orientations, perspective transformations to mimic different viewpoints, and pixel-level transformations to improve generalization in challenging conditions. These techniques are aimed to enhance the model's robustness and accuracy in segmenting catheters under varying alignments, positions, and imaging scenarios.

A few of the augmented images are displayed in Figure 1. This optimal augmentation policy extends the dataset to almost quadruple the size of the training and validation sets (Table 1). The test set remains untouched.

Table 1: Number of images before and after augmentation.

Before Augmentation			After Augmentation		
Training	Validation	Testing	Training	Validation	Testing
1218	87	175	4872	783	175

2.2 Attention-based autoencoder

Since catheters occupy only a few pixels in US images, their representation within the overall image is minimal. To alleviate class-imbalance issues, the study proposes an attention-based convolutional autoencoder to capture the global scene information within an image.

The model architecture is shown in Figure 2. It consists of a U-shaped encoder-decoder structure connected with a bottleneck module. The encoder network on the left side utilizes 2D convolutions with a 3x3 kernel size, followed by ReLU activation to introduce nonlinearity. A structured dropout block[16] with a drop probability of

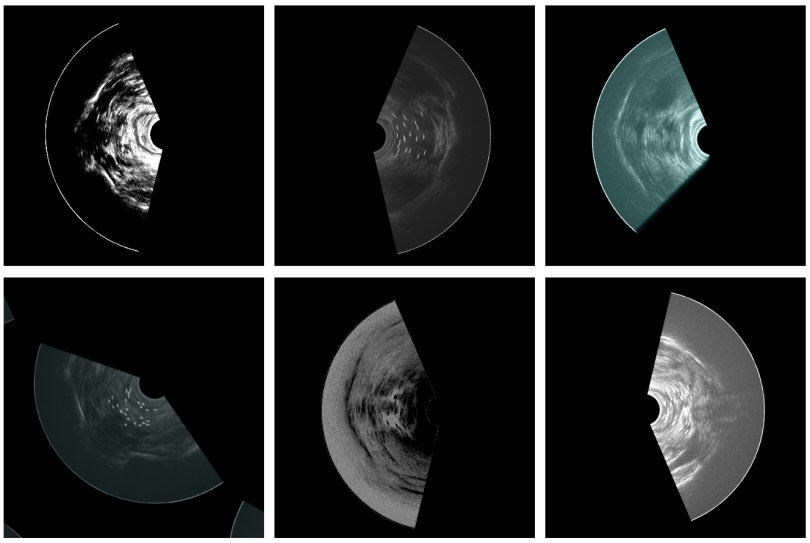


Figure 1: Images generated by our automated augmentation pipeline

0.2 and a window size of 7x7 is included to prevent overfitting. Batch normalization[17] is applied for network regularization. Max-pooling with a 2x2 kernel and stride of two is used for downsampling the feature vector, performed three times, doubling the number of channels each time.
In the decoder section, each block employs a single 2D transposed convolution to reduce the number of channels by half, followed by concatenation with corresponding feature maps from the contracting path. The decoder blocks also consist of two 2D convolutions with a 2x2 kernel and ReLU activation. The output block comprises a 2D convolution layer with one channel, followed by a sigmoid activation layer, generating the required binary segmentation mask.

2.3 Catheter Path Reconstruction From Predicted Catheter Positions

The segmentation model produces 2D masks (i.e. binary images) with predicted catheter locations, where a pixel value of 1 is used to represent a catheter and a pixel value of 0 represents the background. To refine the catheter detection, a contour detection algorithm is applied to identify catheter boundaries. The centroids of these boundaries are then calculated to obtain the center points of each detected catheter, which are subsequently combined to form a 3D point cloud.

Table 2: Modified RANSAC algorithm for Catheter Path Reconstruction

Data: 3D Point Cloud *pts*, Distance Threshold *thresh*, Maximum Iterations *maxIteration*
Result: Line Parameters A, B and Inliers *inliers*

Initialization;
for $it \leftarrow 1$ **to** *maxIteration* **do**
  Randomly select 2 points pt_1, pt_2 from *pts*;
  Compute line vector $vecA = pt_2 - pt_1$;
  Normalize *vecA*;
  Compute distances *dist_pt* from points in *pts* to the line;
  Select inliers *pt_id_inliers* within distance threshold *thresh*;
  if *length(pt_id_inliers) > length(best_inliers)* **then**
    Update *best_inliers*, A, B;

To reconstruct the catheter paths using the generated point cloud, a modified version of the 3D RANSAC algorithm[18] algorithm – shown in Table 2 – is employed. The proposed modified algorithm considers only paths

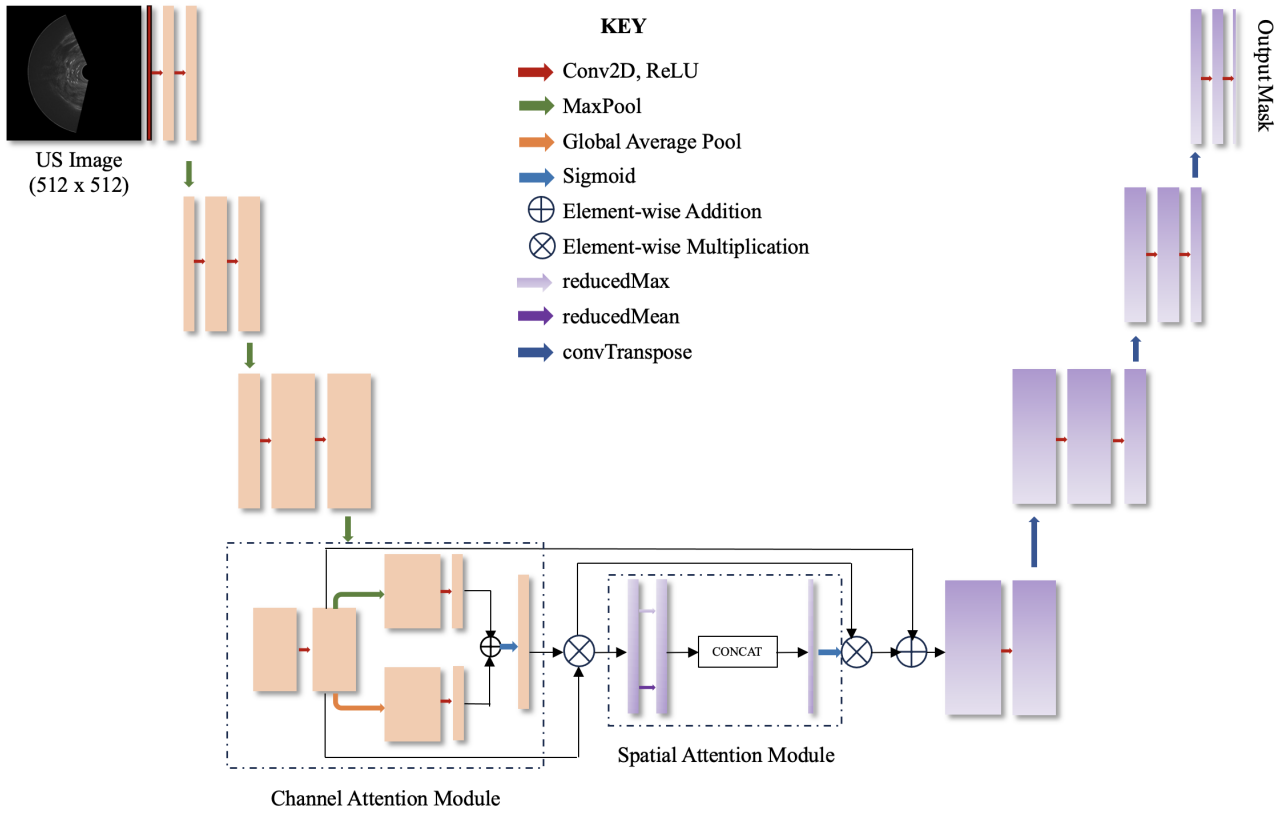


Figure 2: A schematic diagram of the model architecture implemented

with a minimum of 10 points within the threshold. The algorithm iterates until no additional paths can be detected. Once the paths are identified, a simple 3D polyline fitting algorithm is employed to generate the detected catheter paths.

2.4 Model Training

The model is trained from scratch using Kaiming Initialization[19] on a single NVIDI A100 GPU. The Focal Tversky loss function[20] is employed as the chosen loss function, with the parameters alpha (α), beta (β), and gamma (γ) set to 0.7, 0.3, and 1.33, respectively. The model is trained for 100 epochs, utilizing the Adam optimizer with a weight decay of 0.00001. For efficiency, the initial convolutional layer is designed with 16 channels to keep the model compact and fast. The learning rate starts at 0.001 for the first 50 epochs and then decreases to 0.0001 for the following epochs. The final model utilizes a drop-block discard size of 7 with a linear drop rate schedule ranging from 0.05 to 0.25.

2.5 Evaluation

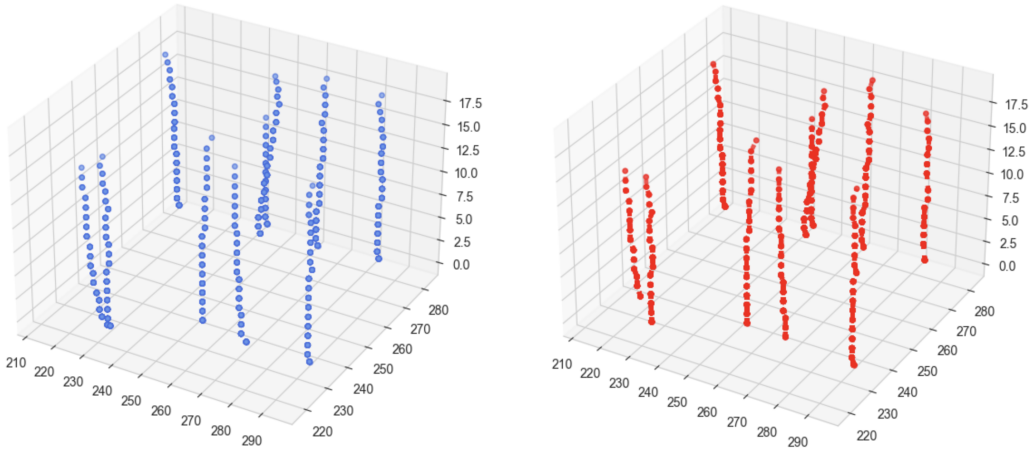
Our method's performance is assessed by comparing it to a U-Net[21] model using three key evaluation metrics: catheter shaft localization error, catheter tip localization error, and catheter detection accuracy.[10,22]
In addition to these metrics, other evaluation measures, namely precision, recall, dice score, intersection over union (IoU) and Matthew's Correlation Coefficient (MCC), were measured. Indeed, Matthew's Correlation Coefficient is appropriate for measurements of performance in binary classification tasks having different data lengths, and can be directly derived from the values for a given confusion matrix itself.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{1}$$

where TP stands for "true positive", FP for "false positive", TN for "true negative" and FN for "false negative". These metrics provide insights into the performance and quality of our method.

3. RESULTS

Our proposed methodology achieved a detection rate of approximately 98% for the tested catheter paths. The mean tip error ranged from 0.067 to 0.311 mm, while the mean shaft error ranged from 0.112 to 0.663 mm, depending on the complexity of the catheter's curve path. Figure 3 illustrates the 3D visualizations of the predicted catheter paths obtained from US images. In Figure 3a, the ground truth is displayed, showing 10 catheters. Figure 3b displays the catheters predicted by our model.



(a) (b)

Figure 3: 3D visualization of the identified catheters on US images - ground truth shown in (a) with our proposed architecture (b).

To further assess performance, we conduct a comparative analysis between our predictions and those obtained from the state-of-the-art medical image segmentation model, U-Net.[21] Figure 4 presents a 2D visualization of predicted catheters on three slices of distinct US images. We observed that our model outperforms U-Net in terms of accurately localizing catheters, as indicated by fewer missed catheter localizations.
Table 3 gives a comprehensive comparison of the evaluation results for different metrics of our model against U-Net architecture and Table 4 shows the detection accuracy, mean-tip error and mean-shaft error for our proposed model architecture and the U-Net architecture. The proposed method increases the detection accuracy compared to the U-Net architecture, and reduces the mean tip error and mean shaft error.

Table 3: A comparison of the evaluation metrics for different methods.

	Dice Score	Precision	Recall	IoU	MCC
U-Net	0.524	0.363	**0.983**	0.361	0.593
Proposed Model	**0.685**	**0.531**	0.982	**0.525**	**0.719**

Table 4: Detection accuracy, tip error and shaft error for different methods.

	Detection Accuracy	Mean Tip Error	Mean Shaft Error
U-Net	93.143	0.283	0.445
Proposed Model	**97.954**	**0.178**	**0.384**

A comparison of inference speed between our proposed model and the U-Net architecture was conducted using a

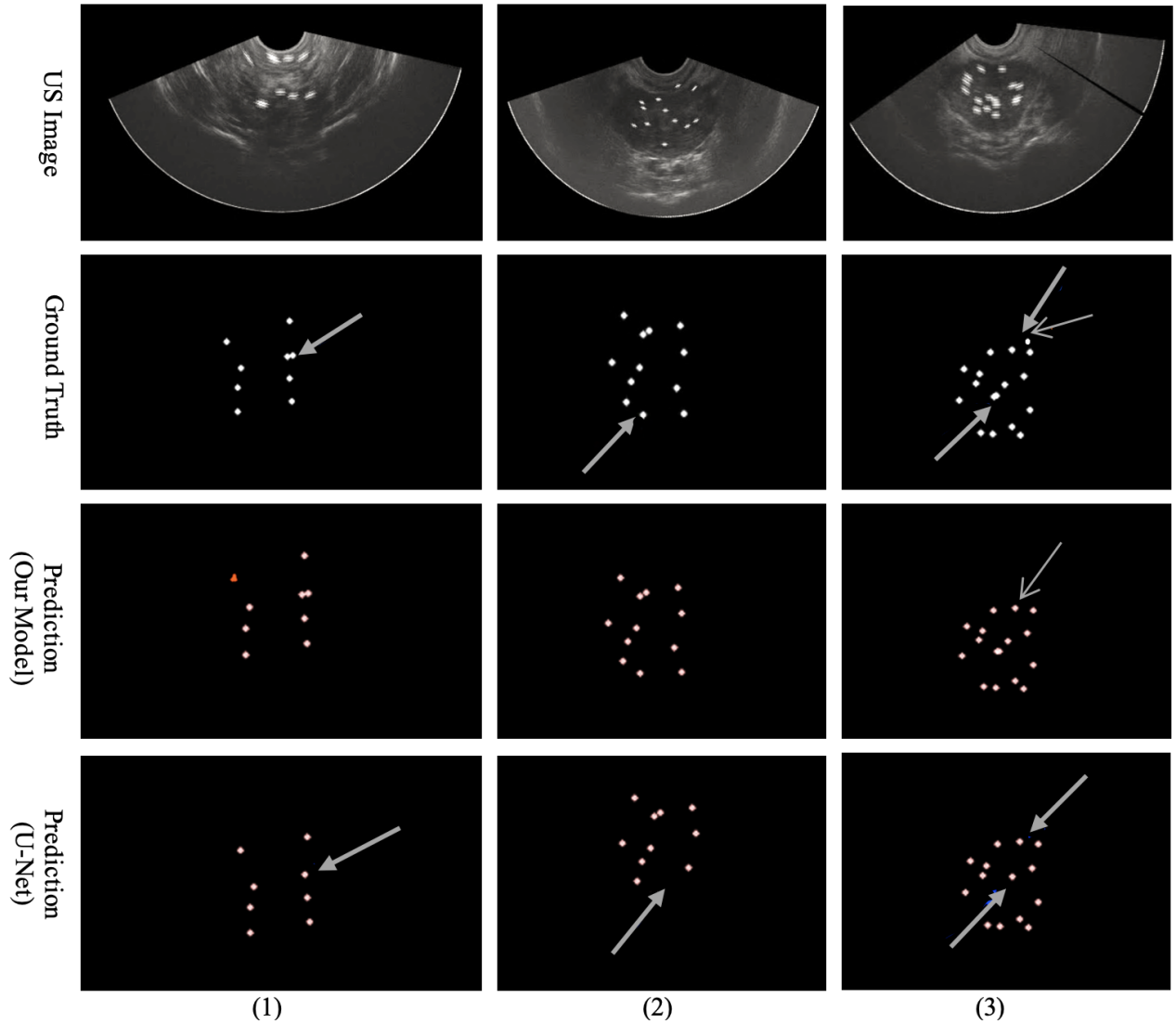


Figure 4: Visualization of 2D slices from various US images. Each column represents a different US image, showcasing the US image itself, the ground truth for catheter localization, predictions made by our proposed method, and predictions from the U-Net architecture.[21] The thick arrows highlight locations where the U-Net architecture missed catheter localizations, while the thin arrows indicate areas where our proposed model missed catheter localizations.

single NVIDIA A100 GPU with 6912 CUDA cores and 40 MB of L2 cache.[23] To mitigate occasional variations, each network was run 10 times, and the average inference speed was calculated at an input resolution of 512 x 512. Our model achieved an average inference speed of 0.0029 seconds, while U-Net had an average inference speed of 0.0057 seconds, highlighting the superior real-time reliability of our model.

4. DISCUSSIONS AND CONCLUSIONS

The proposed pipeline makes significant contributions in several key areas. The attention-based convolutional autoencoder effectively captures global scene information and prioritizes important segments within ultrasound images, resulting in precise catheter segmentation. The 3D RANSAC algorithm ensures robust path reconstruction by accurately identifying catheter boundaries and generating coherent paths. Furthermore, utilizing the

auto-augmentation technique proved valuable in addressing the challenge of limited training data. By optimizing augmentation policies and expanding the dataset, we enhance the model's generalization capabilities, thereby improving its accuracy in reconstructing catheter paths.

While our research demonstrates promising results, there are opportunities for further minimizing tip errors, reducing shaft errors, and decreasing the number of missed catheters. Additionally, applying our pipeline to a more extensive and diverse dataset, including real patient data, would provide further validation of its effectiveness in a clinical setting.

Overall, this research contributes to the advancement of catheter path reconstruction techniques, providing valuable insights and a robust framework to improve treatment planning in HDR brachytherapy and other medical procedures involving catheter tracking.

REFERENCES

[1] Rawla, P., "Epidemiology of prostate cancer," *World journal of oncology* **10**(2), 63 (2019).

[2] Bentzen, J., Ockelmann, H., and Hansen, H., "High dose rate 192ir-microselectron," *Ugeskrift for Laeger* **152**(40), 2908–2910 (1990).

[3] Mendez, L. C. and Morton, G. C., "High dose-rate brachytherapy in the treatment of prostate cancer," *Translational andrology and urology* **7**(3), 357 (2018).

[4] Whitaker, M., Hruby, G., Lovett, A., and Patanjali, N., "Prostate hdr brachytherapy catheter displacement between planning and treatment delivery," *Radiotherapy and oncology* **101**(3), 490–494 (2011).

[5] Poulin, F. and Amiot, L.-P., "Interference during the use of an electromagnetic tracking system under or conditions," *Journal of biomechanics* **35**(6), 733–737 (2002).

[6] Lugez, E., Sadjadi, H., Joshi, C. P., Akl, S. G., and Fichtinger, G., "Improved electromagnetic tracking for catheter path reconstruction with application in high-dose-rate brachytherapy," *International journal of computer assisted radiology and surgery* **12**, 681–689 (2017).

[7] Kindratenko, V. V., "A survey of electromagnetic position tracker calibration techniques," *Virtual Reality* **5**, 169–182 (2000).

[8] Kitner, N., Rodgers, J. R., Ungi, T., Korzeniowski, M., Olding, T., Mousavi, P., and Fichtinger, G., "Multi-catheter modelling in reconstructed 3d transrectal ultrasound images from prostate brachytherapy," in [*Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling*], **12466**, 126–135, SPIE (2023).

[9] Hrinivich, W. T., Hoover, D. A., Surry, K., Edirisinghe, C., Montreuil, J., D'Souza, D., Fenster, A., and Wong, E., "Simultaneous automatic segmentation of multiple needles using 3d ultrasound for high-dose-rate prostate brachytherapy," *Medical physics* **44**(4), 1234–1245 (2017).

[10] Dai, X., Lei, Y., Zhang, Y., Qiu, R. L., Wang, T., Dresser, S. A., Curran, W. J., Patel, P., Liu, T., and Yang, X., "Automatic multi-catheter detection using deeply supervised convolutional neural network in mri-guided hdr prostate brachytherapy," *Medical Physics* **47**(9), 4115–4124 (2020).

[11] Zhang, Y., Lei, Y., Qiu, R. L., Wang, T., Wang, H., Jani, A. B., Curran, W. J., Patel, P., Liu, T., and Yang, X., "Multi-needle localization with attention u-net in us-guided hdr prostate brachytherapy," *Medical physics* **47**(7), 2735–2745 (2020).

[12] Zhao, Y., Cachard, C., and Liebgott, H., "Automatic needle detection and tracking in 3d ultrasound using an roi-based ransac and kalman method," *Ultrasonic imaging* **35**(4), 283–306 (2013).

[13] Fedorov, A., Khallaghi, S., Sánchez, C. A., Lasso, A., Fels, S., Tuncali, K., Sugar, E. N., Kapur, T., Zhang, C., Wells, W., et al., "Open-source image registration for mri–trus fusion-guided prostate interventions," *International journal of computer assisted radiology and surgery* **10**, 925–934 (2015).

[14] Fedorov, A., Nguyen, P. L., Tuncali, K., and Tempany, C., "Annotated MRI and ultrasound volume images of the prostate," (Mar. 2015). Available at https://doi.org/10.5281/zenodo.16396.

[15] Hataya, R., Zdenek, J., Yoshizoe, K., and Nakayama, H., "Faster autoaugment: Learning augmentation strategies using backpropagation," in [*Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*], 1–16, Springer (2020).

[16] Ghiasi, G., Lin, T.-Y., and Le, Q. V., "Dropblock: A regularization method for convolutional networks," *Advances in neural information processing systems* **31** (2018).

[17] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in [*International conference on machine learning*], 448–456, pmlr (2015).

[18] Mariga, L., "pyRANSAC-3D," (10 2022). https://github.com/leomariga/pyRANSAC-3D.

[19] He, K., Zhang, X., Ren, S., and Sun, J., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in [*Proceedings of the IEEE international conference on computer vision*], 1026–1034 (2015).

[20] Abraham, N. and Khan, N. M., "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in [*2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*], 683–687, IEEE (2019).

[21] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*], 234–241, Springer (2015).

[22] Zhang, Y., Tian, Z., Lei, Y., Wang, T., Patel, P., Jani, A. B., Curran, W. J., Liu, T., and Yang, X., "Automatic multi-needle localization in ultrasound images using large margin mask rcnn for ultrasound-guided prostate brachytherapy," *Physics in Medicine & Biology* **65**(20), 205003 (2020).

[23] Choquette, J., Gandhi, W., Giroux, O., Stam, N., and Krashinsky, R., "Nvidia a100 tensor core gpu: Performance and innovation," *IEEE Micro* **41**(2), 29–35 (2021).