2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782105

Toward Autonomous Marker Localization for Lumbar Epidural Steroid Injection Robot

1st Depeng Liu
School of Biomedical Engineering
Shanghai Jiao Tong University
Shanghai, China
liudepeng@sjtu.edu.cn

2nd Ruirui Huang
Department of Computer Science
Johns Hopkins University
Baltimore, USA
rhuang38@jhu.edu

3rd Dimitri Lezcano
Department of Mechanical Engineering
Johns Hopkins University
Baltimore, USA
dlezcan1@jhu.edu

4th Gang Li, Ph.D.
Sheikh Zayed Institute for
Pediatric Surgical Innovation
Children's National Hospital
Washington DC, US
gli2@childrensnational.org

5th Iulian I. Iordachita Ph.D.
Department of Mechanical Engineering
Johns Hopkins University
Baltimore, USA
iordachita@jhu.edu

***Abstract*—This study compares the effectiveness of the traditional minimum circle detection strategy, i.e. Welzl's algorithm, and the state-of-the-art nnU-Net in the localization of lumbar Epidural Steroid Injection (ESI) robot markers across different imaging modalities (MRI and CT). Fiducial frames and markers of identical design were used in both settings. To adjust for human errors in the benchmarking process, experiments were conducted to compare computational and manual marking results. Due to the complexity of the CT dataset, the accuracy and sensitivity of 3D nnU-Net were significantly superior to Welzl's algorithm. However, in the relatively simple MRI datasets, Welzl's algorithm outperformed the learning-based method. Subsequent experiments were conducted to validate that the localization accuracy meets the contingency requirements. This finding informs potential improvement in the current clinical workflow and the registration process of surgical robots in general.**

***Index Terms*—computer vision, segmentation, 3D U-Net, MRI, CT**

I. Introduction

Low back pain ranks among the leading reasons for medical consultations in the United States [1], [2]. Traditional lumbar injections, guided by fluoroscopy or CT, expose patients and physicians to ionizing radiation [3]. In contrast, Magnetic resonance imaging (MRI), non-ionizing and offering high-resolution soft tissue contrast and detailed anatomical visualization, becomes a more ideal imaging modality, albeit with slightly higher costs and longer imaging durations [4], [5].

In recent years, the development of both X-ray and MRI-compatible robots has become notable, most of which rely on either surgeons or their assistants for registration [6]. Typically, this registration process involves manual annotation of at least 25 images, which is laborious and time-consuming [7], [8]. For instance, Monfaredi *et al*. proposed a shoulder-mounted robot for MRI-guided needle placement, where the robot's reference point had to be manually selected and marked on MR images [9]; Maurin *et al*. introduced a patient-mounted robotic platform that required manual marking of the reference point on a graphic interface [10]; Li *et al*. developed a body-mounted robot for lumbar ESI surgeries, which streamlined clinical workflow but still required manual marking of cylindrical fiducial markers [11], [12]. The accuracy of these registrations also depended highly on the expertise of the surgeons, as noted by Machetanz *et al* [13].

Previous efforts have aimed to automate this registration process. Krigger *et al*. proposed active and passive methods for registering cylindrical fiducial markers in MRI settings [14]; Tokuda *et al*. introduced the Z-shaped frame marker technique, which eliminates the necessity of manually selecting cylindrical markers [15]. However, both of these methods involved complex preparation and required special materials, making them challenging to transfer to ESI robots.

To tackle this challenge, we attempted and successfully employed traditional and Deep Learning (DL) methods to segment fiducial markers from their background in CT and MR images. The key contributions of our study are as follows: (1) Applied Welzl's algorithm and nnU-Net to autonomous marker localization in lumbar ESI robots, with independent tuning and evaluation on both MRI and CT imaging modalities; (2) Designed a unified robotic fiducial frame for CT and MRI so conclusions can be drawn across methods and modalities; (3) Analyzed the pros and cons of traditional and DL-based methods across different modalities.

II. Materials and Methods

A. Experimental Setup

As depicted in Figure 1, the fiducial frame was designed identically for both CT and MRI, but fabricated using different materials to ensure compatibility and visibility within the respective modality during the localization process. In CT imaging, the fiducial frame was 3D-printed using Acrylonitrile Butadiene Styrene(ABS) material (STRATASYS, Eden

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.

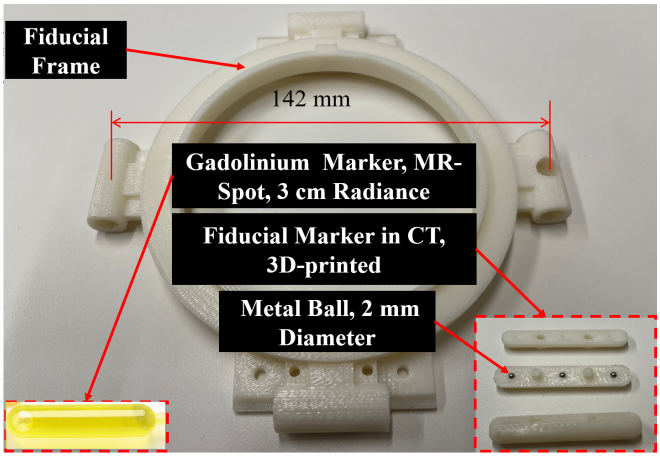

Fig. 1. Fiducial frame and fiducial markers in both CT and MRI.

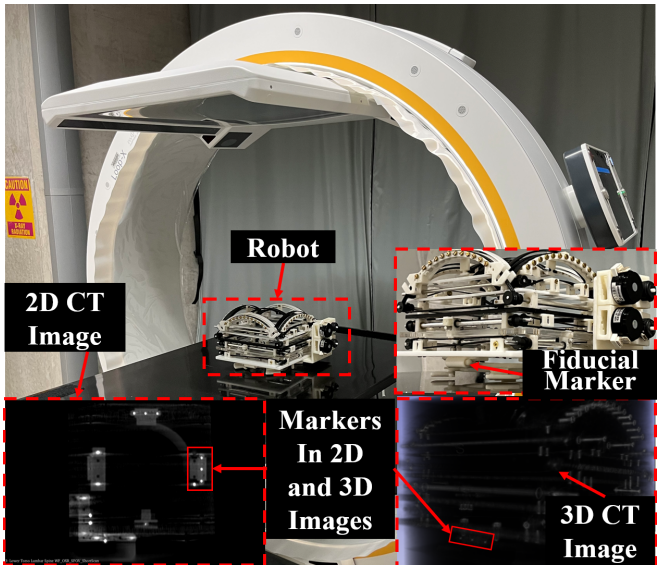

Fig. 2. Epidural Steroid Injection robots with fiducial frame in CT images. 2D CT images with fiducial markers(Left). Real robot with fiducial frame in CT scanning and 3D CT images of the fiducial markers(Right).

Prairie, USA), a radiopaque material that does not interfere with the imaging process. Within the ABS frame, three 2 mm-diameter metal balls were embedded as identification targets, which offer high contrast and visibility on CT images. In contrast, for MR imaging, the fiducial markers were made from Gadolinium material (specifically, MR-Spot 3.0 cm Radiance) since their paramagnetic properties provide clear and distinct signal intensity on MRI scans and enable precise localization in lumbar ESI procedures.

Figure 2 illustrates the robot structure and the relative positioning of the fiducial frame. Within the frame, four cylindrical fiducial markers were embedded: three horizontally placed and one orthogonal to the rest for registration purposes. Each cylindrical fiducial marker contains three metal balls, spaced 11 mm apart. The positions of the middle balls were used as reference points during registration. The ground truth for the segmentation of metal balls and cylindrical fiducial markers was obtained through manual annotations. According to previous research, the error in the manually labeled registration process is 1.00 ± 0.73 mm [16]. Therefore, the error requirement for automatic labeling is set to be 1 mm.

B. Welzl's Algorithm

Welzl's algorithm, proposed by Emo Welzl in 1991 [17], is a simple yet effective randomized algorithm for the minimum covering circle problem that runs in expected O(n) time. Since metal balls and fiducial markers appear either circular or elliptical under CT and MR scans, Welzl's algorithm was deemed a robust algorithm for marker segmentation.

The correctness of Welzl's algorithm is based on the simple fact that: (1) The MEC is unique, and (2) such a circle can be determined by at most three points that lie on the boundary of the circle. Starting with a set of points, Welzl's algorithm randomly selects one point and recursively finds the minimal circle containing the remaining points, If the returned circle also encloses the point, it is the minimal circle and is returned. Otherwise, this point must lie on the boundary of the resulting circle. The algorithm recurses on a set of points known to be on the circle boundary and terminates when either an empty set is reached, or when 3 points remain, at which the circle is the circum-circle of the triangle described by the points. If we denote the coordinates of the circle's center as (x, y) and the radius of the circle as r, then for every point (x_i, y_i) within the circle, the following relation

$$(x_i - x)^2 + (y_i - y)^2 \leq r^2, \tag{1}$$

which ensures that all points (x_i, y_i) lie within the circle. Our goal is to find the smallest possible r such that this condition is satisfied for all points.

Before applying Welzl's algorithm, Gaussian blur filtering is first applied to the image, followed by the use of the canny edge detector to identify the image's edges. In three-dimensional space, the CT image projection of a metal ball forms a circular shape, irrespective of the scanning direction. In the cross-sectional representations of a fiducial marker, as observed in MRI, the anatomical configuration encompasses one circular and three entities of elliptical shape. By assessing each boundary point, potential circle and ellipse centers are detected.

C. Architecture of the nnU-Net

A six-stage 3D U-net architecture with high resolution was employed [18], featuring a classic Encoder-decoder design with skip connections. The model, illustrated in Figure 3, utilized a batch size of two and a patch size of $80 \times 128 \times 192$, a sliding window with half-patch overlap upon each stride, Dice and cross-entropy loss function, and a Poly learning rate (0.01) scheduler. Each stage comprised two convolution layers. The network was trained for 350 epochs.

The initial model configuration was facilitated by nnU-Net, an end-to-end, DL-based, task-agnostic segmentation tool

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.

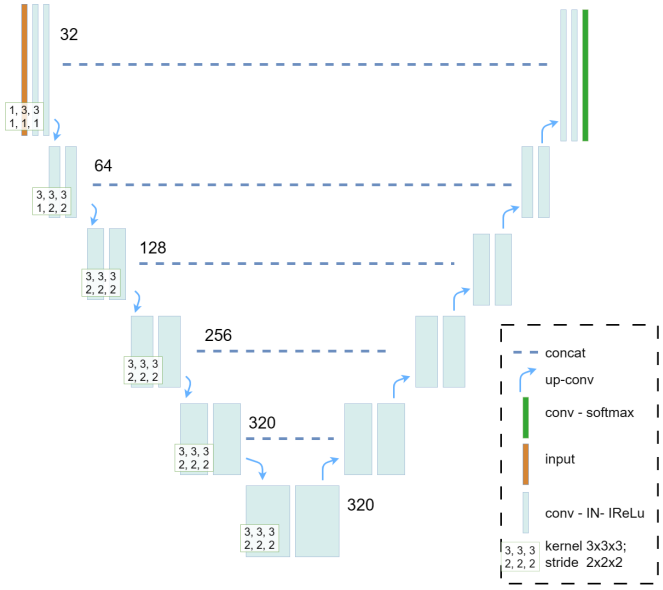

Fig. 3. 3D nnU-Net architectural diagram. Skip connections capture global contexts and improve segmentation results.

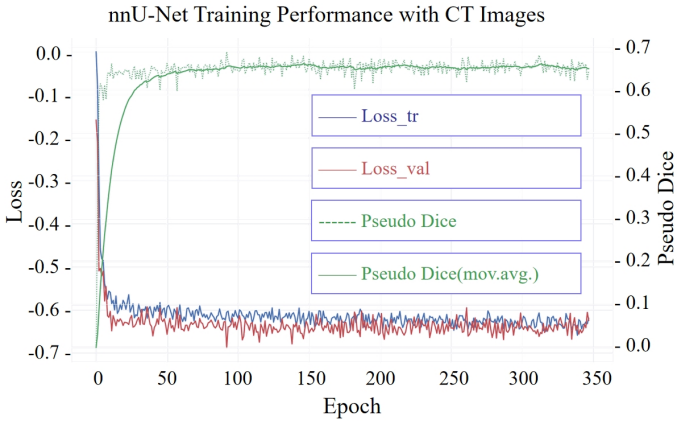

Fig. 4. Training process of the nnU-Net method with validation loss curve in CT images.

of large pre-trained data. Model parameters were divided into three categories: "fixed", "rule-based", or "empirical". These parameters were either set universally or dynamically determined during the training stage based on the intrinsic characteristics of features extracted from the input dataset.

Upon completion, nnU-Net outputs the tuned model parameters in the form of a JSON file, which serves as a blueprint to fully re-construction the entire network architecture, and was a starting point for manual fine-tuning and modifications.

III. Experiment and Results

A. CT and MRI Dataset

CT and MR images were independently collected. The MRI dataset consisted of images covering coronal, sagittal, and axial planes obtained from 8 different scans, each conducted at a distinct robotic angle. These images were acquired using a 1.5T MRI scanner (Aera, Siemens, Germany) with an original pixel resolution of 288×384 and 620 images in total. The CT dataset included 38 fully scanned robot scenes captured by the Loop-X (Brainlab, Munich, Germany) in various positions and postures, with a pixel size of 1385×1385×1386 and a spatial resolution of 0.2 mm.

B. Workflow and Experiment

CT images used to train nnU-Net were split into training, and testing datasets at a ratio of 30: 8. The dataset was split into three categories for training cases, test cases, and training masks as shown in Figure 3. The pre-processor read the image files and generated a JSON file containing the extracted metadata. The model was trained using a computational system with an RTX 4070 GPU.

The training process of the nnU-Net with CT images is shown in Figure 4. As training progressed, the cross-entropy loss decreased as the model learned to extract representative features from the training data. The decreasing validation loss indicates that the model generalizes well to unseen data without over-fitting. The pseudo-dice coefficient measures the overlap between predicted and ground truth segmentation, which confirms that the model has improved over time in segmenting the object of interest.

MR images for training nnU-Net were initially stored as 2D DICOM files, then reconstructed using 3D Slicer to generate three-dimensional representations of the imaged objects. Fiducial markers were manually segmented and served as the ground truth for subsequent model training and evaluation. Both the raw images and segmentation masks were converted into the binary NRRD format. Due to the limited dataset size, training and testing sets were split evenly. The other procedures are identical to those employed in CT processing.

The workflow for Welzl's algorithm dealing with MR images includes pre-processing, MEC detection, and post-

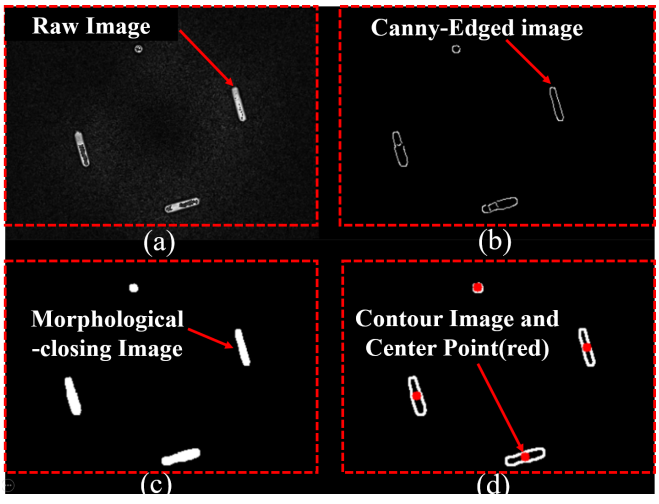

Fig. 5. Procedure of the Welzl's algorithm and the results on MR images. (a) Raw image. (b) Canny-Edged image. (c) Morphological-closing Image. (d) Contour image and center point(red).

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.

processing. The preprocessing step included applying Gaussian blur, Canny edge detection, and morphological closing to the raw image to enhance contours. Welzl's algorithm was followed by a post-processing step that computes the centers of these circles for quantitative analysis. The process is visually depicted in Figure 5, illustrating the progression from the original MR image to the final image annotated with the detected feature centers. Welzl's algorithm worked successfully for MR images but failed to correctly segment CT images due to a variety of reasons elaborated later in the Discussion section.

C. Marking Error Evaluation

The manual marking error is defined as the distance between the centroid of the fiducial markers and manually labeled points. In manual marking, the error is calculated through 30 repetitions of the same labeling action. The definition of MRI error within Welzl's algorithm is expressed as the average distance between all detected two-dimensional coordinate positions and the centroid of the two-dimensional image. For nnU-Net, MRI error is defined as the average distance between the centroids in the training set and the test set and CT error is defined as the mean difference in the distances between the labeled marker points and their corresponding centroids in the training and test sets.

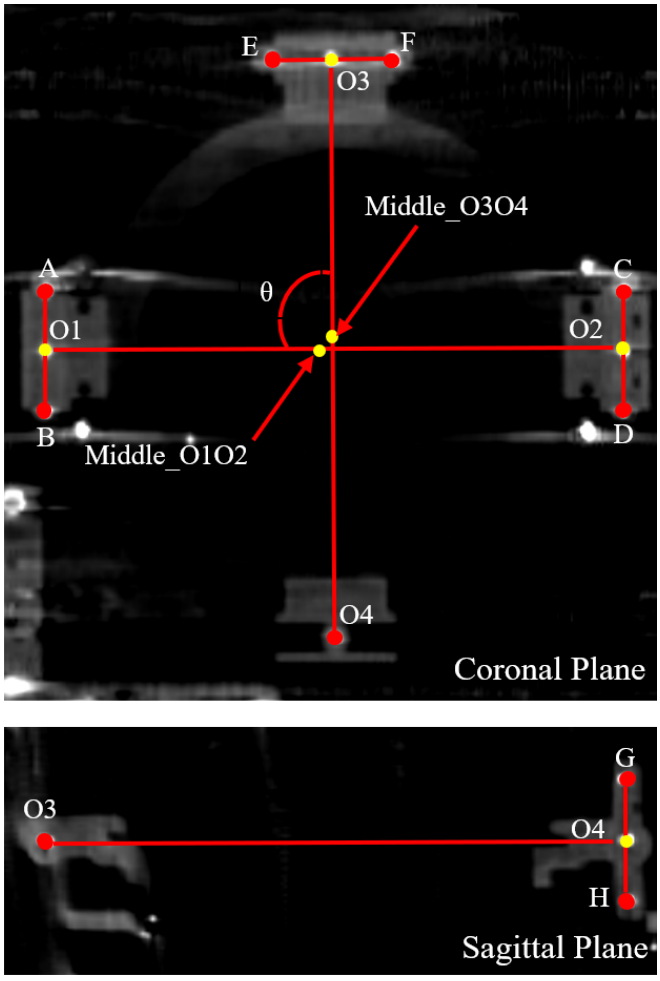

Fig. 6. Schematic illustration of fiducial frame verification parameters on the coronal plane and sagittal plane of CT images.

TABLE I
DETECTION ERROR EVALUATION ON CT AND MRI.

Method	CT (STD)	Method	MRI (STD)
Welzl's algorithm	N/A	Welzl's algorithm	0.73 ± 0.31
nnU-Net	0.86 ± 0.44	nnU-Net	1.15 ± 0.83
Manual	0.63 ± 0.18	Manual	0.44 ± 0.43

Unit: mm.

For CT scans with the nnU-Net, the average marking error is 0.86 mm with a standard deviation of 0.44 mm, higher than Welzl's algorithm and comparable to the manual marking error 0.63 $\pm$ 0.18 mm, as shown in Table I.

For MR images, nnU-Net's accuracy is 1.15 $\pm$ 0.83 mm, slightly below its counterpart Welzl's algorithm, which has an accuracy of 0.73 $\pm$ 0.31 mm. the manual marking error 0.44 $\pm$ 0.43 mm is lower than both of them. Results are illustrated in Table I.

D. Position Error Evaluation on CT images

The following parameters were defined based on the structure of the fiducial frame in CT images for the most accurate benchmarking result. As shown in Figure 6, six points were defined in the horizontal direction as A to F, and two points in the vertical direction as G and H. The distance between the midpoint of line AB(O1) and the midpoint of line CD(O2) is defined as l_1. The distance between the midpoint of line EF(O3) and the midpoint of line GH(O4) is defined as l_2. The distance between the midpoint of line O1O2(Middle_O1O2) and the midpoint of line O3O4(Middle_O3O4) is defined as l_3. The angle θ is between the projection O1O2 and O3O4 on the coronal plane.

The actual values for l_1, l_2, l_3, and θ are determined as 142 mm, 142 mm, 0 mm, and 90°, respectively. Figure 7 displays the distribution of errors between these measurements and their respective nominal values.

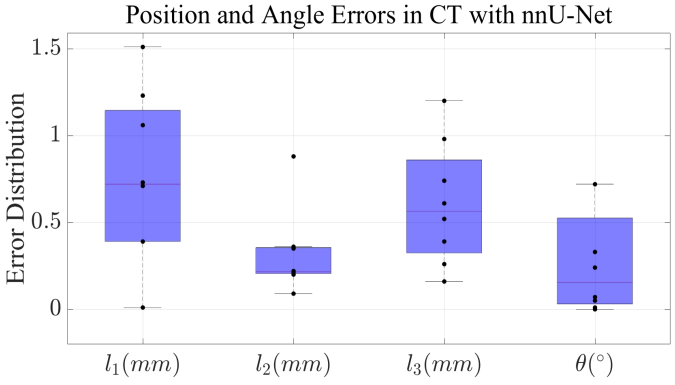

Fig. 7. Error distribution of the verification parameters on CT images.

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.

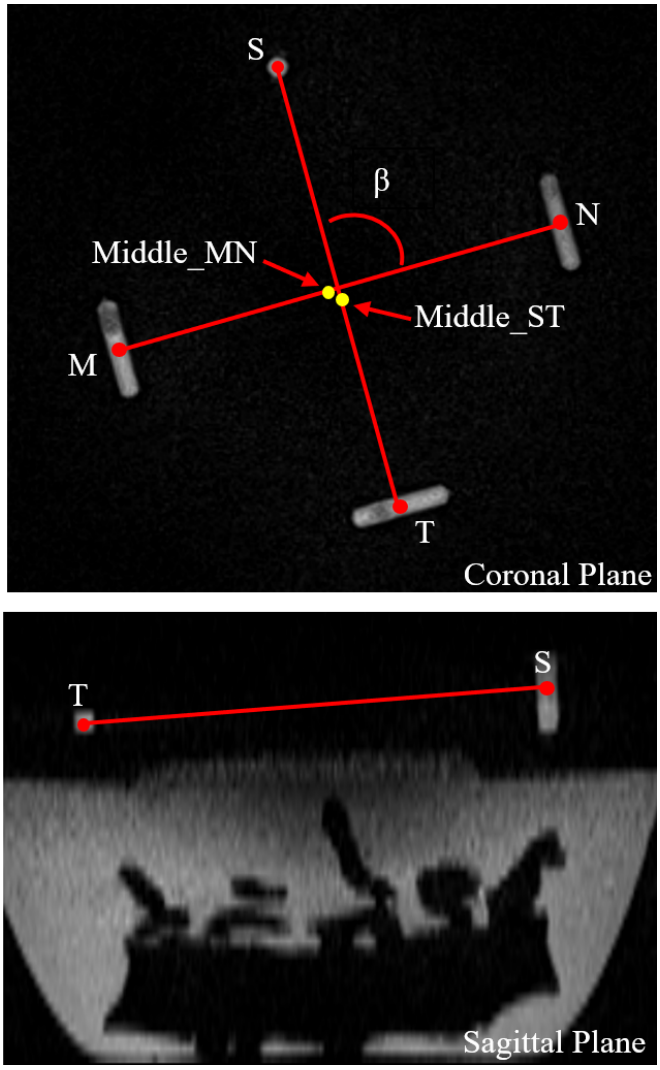

Fig. 8. Schematic illustration of fiducial frame verification parameters on the coronal plane and sagittal plane of MRI.

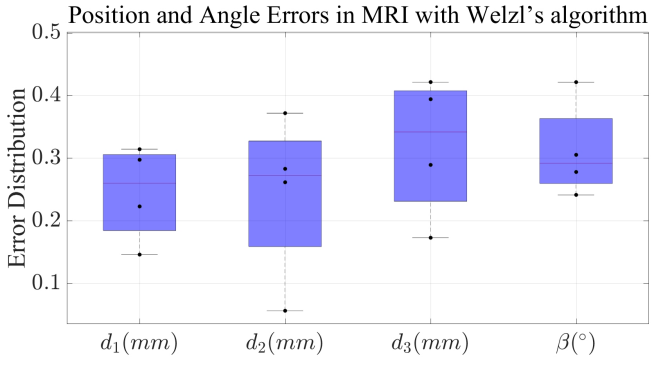

Fig. 9. Error distribution of the verification parameters on MRI.

E. Position Error Evaluation on MRI

As visualized in Figure 8, we defined the following parameters for position verification: Four points M, N, S, and T are the marker centers. d_1, d_2, and d_3 are verification distances between M and N, S and T, and the midpoint of MN(Middle_MN) and the midpoint of ST(Middle_ST), respectively. verification angle β is between projection MN and ST on the coronal plane.

Based on the model information, the theoretical values of d_1, d_2, d_3, and β should be 142 mm, 142 mm, 0 mm, and 90°, respectively. We measured the actual values of these four parameters in the MRI results and calculated the distribution of the errors between the actual values and the theoretical values as shown in Figure 9.

IV. Discussion and Future Work

Welzl's algorithm, with its linear asymptotic time efficiency, is highly efficient for computing the minimum enclosing circle of a given set of points, presenting a significant advantage when handling large datasets. Its probabilistic recursive approach mitigates worst-case scenarios common in deterministic algorithms.

For MR images, characterized by the markers' relatively simple shape and fewer noise artifacts, Welzls's algorithm outperformed nnU-Net with high accuracy and sensitivity, achieving a mean accuracy of 0.73 $\pm$ 0.31 mm. The learning-based method offers slightly lower accuracy compared to traditional computer vision methods, with the verification distance d_1, d_2, d_3, and verification angle β less than 0.25 mm, 0.24 mm, 0.32 mm, and 0.31°, respectively as shown in Figure 9. This result could be attributed to potential errors introduced during the conversion of MRI images into three-dimensional data, a process that involved the integration of manually annotated markers. These errors were likely amplified during the subsequent learning phase within the nnU-Net framework. Since the requirement of the accuracy error is around 1 mm, the accuracy of 1.15 $\pm$ 0.83 mm is still considered acceptable.

For CT images, Welzl's algorithm failed to segment markers due to the visibility of the robot under CT scans. The spherical parts within the robot structure introduce noise, complicating the detection process and increasing the false positive rate, as highlighted in Figure 2. Tn this case, 3D-UNet model was able to successfully segment CT images with a mean error of 0.86 $\pm$ 0.44 mm. The mean error for verification distances l_1, l_2, and l_3, were all less than 0.75 mm, 0.32 mm, and 0.36 mm, respectively; The mean error for the angle θ was within 0.19°.

The complexity of the CT scans naturally lends itself to Deep Learning-based strategies. Through extensive testing, We ultimately went with the 3D high-resolution nnU-Net architecture due to a few considerations: The first problem is the scarcity of medical imaging data for training deep learning models, which could be solved by nnU-Net's large and diverse pool of pre-trained data that is highly translational to unseen images. The second problem is data characteristics. The metal balls represented only 1:24,000 of the total volume in a 3D CT image, and the cylindrical markers represented only 0.5% of the total volume in a 3D reconstructed MR image. Concerns initially arose regarding convolutional neural networks' ability to capture such characters in high-level features accurately. The method nnU-Net with the iterative self-

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.

configuring process that systematized the empirical turning process to automatically determine optimal parameters through rule-based embedding, fingerprint extraction, and pipeline extraction, minimizes the process of trials and errors to find the suitable architecture and configuration. The last concern was the need to preserve 3D features, which is also solved by nnU-Net since it provides one 2D and two 3D U-Net training options.

In the future, we plan to enhance the robustness of the model by adopting transfer learning strategies and expanding the scope of our medical image dataset. We plan to conduct extensive benchmarking experiments aimed at refining the registration accuracy of lumbar Epidural Steroid Injection robots.

V. Conclusion

In this paper, we propose different algorithms capable of automatically recognizing different fiducial markers of lumbar ESI robots in CT and MR images. Preliminary results indicate that employing classic and learning-based algorithms can effectively reduce the labor-intensive nature of manual annotation by clinicians with comparable accuracy.

VI. Acknowledgment

This work was primarily supported by the NIH through Grants 1R01 EB025179 and the Multi-Scale Medical Robotics Center in Hong Kong. The author Depeng Liu would like to express his appreciation to the National Key Research and Development Program of the Ministry of Science and Technology 2022YFC2406900, National Natural Science Foundation of China 81727806, and China Scholarship Council 202106230154 for their valuable support.

References

[1] Ivan Urits, Aaron Burshtein, Medha Sharma, Lauren Testa, Peter A Gold, Vwaire Orhurhu, Omar Viswanath, Mark R Jones, Moises A Sidransky, Boris Spektor, et al. Low back pain, a comprehensive review: pathophysiology, diagnosis, and treatment. *Current pain and headache reports*, 23:1–10, 2019.

[2] Simon Dagenais, Jaime Caro, and Scott Haldeman. A systematic review of low back pain cost of illness studies in the united states and internationally. *The spine journal*, 8(1):8–20, 2008.

[3] Umile Giuseppe Longo, Mattia Loppini, Luca Denaro, Nicola Maffulli, and Vincenzo Denaro. Rating scales for low back pain. *British medical bulletin*, 94(1):81–144, 2010.

[4] John Bui and Nikolai Bogduk. A systematic review of the effectiveness of ct-guided, lumbar transforaminal injection of steroids. *Pain Medicine*, 14(12):1860–1865, 2013.

[5] Kevin A Smith and John Carrino. Mri-guided interventions of the musculoskeletal system. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(2):339–346, 2008.

[6] Hao Su, Ka-Wai Kwok, Kevin Cleary, Iulian Iordachita, M. Cenk Cavusoglu, Jaydev P. Desai, and Gregory S. Fischer. State of the art and future opportunities in mri-guided robot-assisted surgery and interventions. *Proceedings of the IEEE*, 110(7):968–992, 2022.

[7] Gang Li, Niravkumar A Patel, Jan Hagemeister, Jiawen Yan, Di Wu, Karun Sharma, Kevin Cleary, and Iulian Iordachita. Body-mounted robotic assistant for mri-guided low back pain injection. *International journal of computer assisted radiology and surgery*, 15(2):321–331, 2020.

[8] Depeng Liu, Gang Li, Shuyuan Wang, Zixuan Liu, Yanzhou Wang, Laura Connolly, David E Usevitch, Guofeng Shen, Kevin Cleary, and Iulian Iordachita. A magnetic resonance conditional robot for lumbar spinal injection: Development and preliminary validation. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 20(1):e2618, 2024.

[9] Reza Monfaredi, Iulian Iordachita, Emmanuel Wilson, Raymond Sze, Karun Sharma, Axel Krieger, Stanley Fricke, and Kevin Cleary. Development of a shoulder-mounted robot for mri-guided needle placement: phantom study. *International journal of computer assisted radiology and surgery*, 13:1829–1841, 2018.

[10] Benjamin Maurin, Bernard Bayle, Olivier Piccin, Jacques Gangloff, Michel de Mathelin, Christophe Doignon, Philippe Zanne, and Afshin Gangi. A patient-mounted robotic platform for ct-scan guided procedures. *IEEE Transactions on Biomedical Engineering*, 55(10):2417–2425, 2008.

[11] Gang Li, Niravkumar A Patel, Weiqiang Liu, Di Wu, Karun Sharma, Kevin Cleary, Jan Fritz, and Iulian Iordachita. A fully actuated body-mounted robotic assistant for mri-guided low back pain injection. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5495–5501. IEEE, 2020.

[12] Gang Li, Niravkumar A Patel, Yanzhou Wang, Charles Dumoulin, Wolfgang Loew, Olivia Loparo, Katherine Schneider, Karun Sharma, Kevin Cleary, Jan Fritz, et al. Fully actuated body-mounted robotic system for mri-guided lower back pain injections: Initial phantom and cadaver studies. *IEEE robotics and automation letters*, 5(4):5245–5251, 2020.

[13] Kathrin Machetanz, Florian Grimm, Martin Schuhmann, Marcos Tatagiba, Alireza Gharabaghi, and Georgios Naros. Time efficiency in stereotactic robot-assisted surgery: an appraisal of the surgical procedure and surgeon's learning curve. *Stereotactic and Functional Neurosurgery*, 99(1):25–33, 2021.

[14] Axel Krieger, Robert C Susil, Cynthia Ménard, Jonathan A Coleman, Gabor Fichtinger, Ergin Atalar, and Louis L Whitcomb. Design of a novel mri compatible manipulator for image guided prostate interventions. *IEEE Transactions on Biomedical Engineering*, 52(2):306–313, 2005.

[15] Junichi Tokuda, Gregory S Fischer, Simon P DiMaio, David G Gobbi, Csaba Csoma, Philip W Mewes, Gabor Fichtinger, Clare M Tempany, and Nobuhiko Hata. Integrated navigation and control software system for mri-guided robotic prostate interventions. *Computerized Medical Imaging and Graphics*, 34(1):3–8, 2010.

[16] Junichi Tokuda, Sang-Eun Song, Kemal Tuncali, Clare Tempany, and Nobuhiko Hata. Configurable automatic detection and registration of fiducial frames for device-to-image registration in mri-guided prostate interventions. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part III 16*, pages 355–362. Springer, 2013.

[17] Emo Welzl. Smallest enclosing disks (balls and ellipsoids). In *New Results and New Trends in Computer Science: Graz, Austria, June 20–21, 1991 Proceedings*, pages 359–370. Springer, 2005.

[18] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Authorized licensed use limited to: Johns Hopkins University. Downloaded on November 28,2025 at 19:49:56 UTC from IEEE Xplore. Restrictions apply.