



Deep learning-based automatic pipeline for 3D needle localization on intra-procedural 3D MRI

Wenqi Zhou^{1,2} · Xinzhou Li^{1,2} · Fatemeh Zabihollahy^{1,3} · David S. Lu¹ · Holden H. Wu^{1,2}

Received: 19 October 2023 / Accepted: 9 February 2024
© The Author(s) 2024

Abstract

Purpose Accurate and rapid needle localization on 3D magnetic resonance imaging (MRI) is critical for MRI-guided percutaneous interventions. The current workflow requires manual needle localization on 3D MRI, which is time-consuming and cumbersome. Automatic methods using 2D deep learning networks for needle segmentation require manual image plane localization, while 3D networks are challenged by the need for sufficient training datasets. This work aimed to develop an automatic deep learning-based pipeline for accurate and rapid 3D needle localization on in vivo intra-procedural 3D MRI using a limited training dataset.

Methods The proposed automatic pipeline adopted Shifted Window (Swin) Transformers and employed a coarse-to-fine segmentation strategy: (1) initial 3D needle feature segmentation with 3D Swin UNet Transformer (UNETR); (2) generation of a 2D reformatted image containing the needle feature; (3) fine 2D needle feature segmentation with 2D Swin Transformer and calculation of 3D needle tip position and axis orientation. Pre-training and data augmentation were performed to improve network training. The pipeline was evaluated via cross-validation with 49 in vivo intra-procedural 3D MR images from preclinical pig experiments. The needle tip and axis localization errors were compared with human intra-reader variation using the Wilcoxon signed rank test, with $p < 0.05$ considered significant.

Results The average end-to-end computational time for the pipeline was 6 s per 3D volume. The median Dice scores of the 3D Swin UNETR and 2D Swin Transformer in the pipeline were 0.80 and 0.93, respectively. The median 3D needle tip and axis localization errors were 1.48 mm (1.09 pixels) and 0.98° , respectively. Needle tip localization errors were significantly smaller than human intra-reader variation (median 1.70 mm; $p < 0.01$).

Conclusion The proposed automatic pipeline achieved rapid pixel-level 3D needle localization on intra-procedural 3D MRI without requiring a large 3D training dataset and has the potential to assist MRI-guided percutaneous interventions.

Keywords Interventional MRI · Device tracking · Transformer networks · Deep learning · Image segmentation · Needle feature

Introduction

Image-guided percutaneous interventions play key roles in cancer diagnosis and treatment with their minimally invasive

characteristics [1, 2]. Currently, most percutaneous interventions are guided by ultrasound (US) or computed tomography (CT) [3, 4]. However, US and CT suffer from insufficient soft-tissue contrast and poor visibility of several important classes of tumors [5, 6]. In comparison, magnetic resonance imaging (MRI) provides excellent soft-tissue contrast and can be the only modality for visualizing tumors that are not visible on CT or US, which makes it an emerging imaging modality for guiding percutaneous interventions in various applications including needle-based targeted biopsy and focal ablation in the liver or other abdominal organs [7–11].

Despite the advantages of MRI-guided interventions, accurate and rapid 3D localization of the interventional needle in intra-procedural MR images remains a major

✉ Holden H. Wu
HoldenWu@mednet.ucla.edu

¹ Department of Radiological Sciences, University of California Los Angeles, 300 UCLA Medical Plaza, Suite B119, Los Angeles, CA 90095, USA

² Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, USA

³ Joint Department of Medical Imaging, Sinai Health System and University of Toronto, Toronto, Canada

challenge [12, 13]. The interventional needle can be visualized on MR images based on the passive signal void feature caused by needle-induced susceptibility effects [14]. In current workflows, 3D needle localization is performed manually; interventional radiologists locate the needle by marking the needle entry point and tip on the intra-procedural 3D MR images [15]. However, manual needle localization requires expert knowledge and is time-consuming, which leads to cumbersome workflows, prolonged procedure time, and potential variability [16, 17]. The lack of timely feedback regarding needle and target locations also hinders the possibility of real-time MRI-guided interventions under human operation or with robotic assistance [18, 19].

2D deep learning networks for automatic 2D needle segmentation and localization have shown promising results [20–22]. However, 2D needle localization methods commonly require the initial manual localization of a 2D image plane that contains the needle feature [23]. Moreover, the 2D needle localization results do not immediately provide the essential 3D relative positions of the needle and the target needed for guiding needle insertion. On the other hand, training 3D deep learning networks for needle segmentation on MR images typically requires large 3D training datasets [24], which may not be available for specific MRI-guided procedures or at specific facilities. Studies of applying 3D deep learning networks for needle segmentation on CT and US images have similarly demonstrated the data-demanding nature of these networks [25, 26]. The potentially limited sizes of intra-procedural 3D MR image datasets and the variabilities in the needle feature's location and grayscale appearance in *in vivo* 3D MRI may lead to insufficient training of the 3D deep learning network and result in inaccurate 3D needle feature segmentation and localization.

For the task of 3D needle segmentation, which requires delineating a relatively small object in a large field-of-view (FOV), convolutional neural network (CNN)-based 3D neural networks may be suboptimal since the convolution operations lack the ability to efficiently capture global information [27, 28]. To better model the long-range information in large FOVs, researchers have developed transformer-based networks that adopt the self-attention mechanism to capture global interactions between contexts [29]. The Shifted Window (Swin) Transformer introduced by Liu et al. demonstrated excellent results with its hierarchical architecture, which enables the model to capture both local and global information [30]. In the context of 3D medical image segmentation, Hatamizadeh et al. further introduced the 3D Swin UNet Transformer (UNETR) [31] which utilized a U-shaped network structure with a Swin Transformer-based encoder and CNN-based decoder. Researchers demonstrated the efficacy of Swin UNETR in 3D medical image segmentation in the BraTS 2021 segmentation challenge, where it outperformed UNet and nnU-Net [31].

In this work, our objective was to develop an automatic pipeline for rapid and accurate 3D needle localization on 3D MRI by taking advantage of transformer networks. To overcome the restriction of limited 3D datasets for training, we combined the 3D Swin UNETR and 2D Swin Transformer for coarse-to-fine segmentation and adopted pre-training and data augmentation strategies. The proposed pipeline was evaluated using *in vivo* 3D MR images acquired during MRI-guided liver interventions in preclinical pig models and compared with manual localization of the 3D needle feature.

Methods

MRI-guided interventional experiments

In an animal research committee-approved study, we performed MRI-guided targeted needle placement in the livers of seven healthy female pigs on a 3 T scanner (MAGNETOM Prisma, Siemens, Erlangen, Germany). These experiments were designed and performed by an experienced interventional radiologist (over 20 years of experience) based on step-and-shoot workflows that mimic clinical image-guided procedures at our institution [32–36].

The workflow of the experiments is shown in Fig. 1. In the planning stage, preoperative 3D T₁-weighted (T1w) gradient echo (GRE) Dixon MR images were acquired to localize the target and initialize the needle entry point and trajectory. In the insertion and confirmation stage, manual needle localization was performed by marking the needle entry point and needle tip on the 3D T1w GRE images in a graphical environment (3D Slicer) [37]. The patient table was moved out from the scanner for the interventional radiologist to insert and adjust the needle based on the 3D relative position of the needle tip and the target determined from MRI. This process was repeated until the needle tip reached the target.

Intra-procedural MRI datasets

Intra-procedural 3D T1w GRE Dixon MR images containing the needle feature and 2D real-time golden-angle (GA) ordered radial GRE images with the image plane aligned with the needle axis were collected during experiments with parameters shown in Table 1. In each experiment, 7 3D T1w GRE images were acquired as confirmation images between needle adjustment steps, with the needle inserted at different depths and angles. Based on the needle location in the 3D confirmation images, 2D real-time GA ordered radial GRE images were acquired on manually located 2D oblique axial and sagittal planes aligned with the needle axis. In each experiment, 70 2D image frames with different insertion depths and angles were selected from the multiple real-time scans to form the 2D radial GRE dataset. Under the guidance

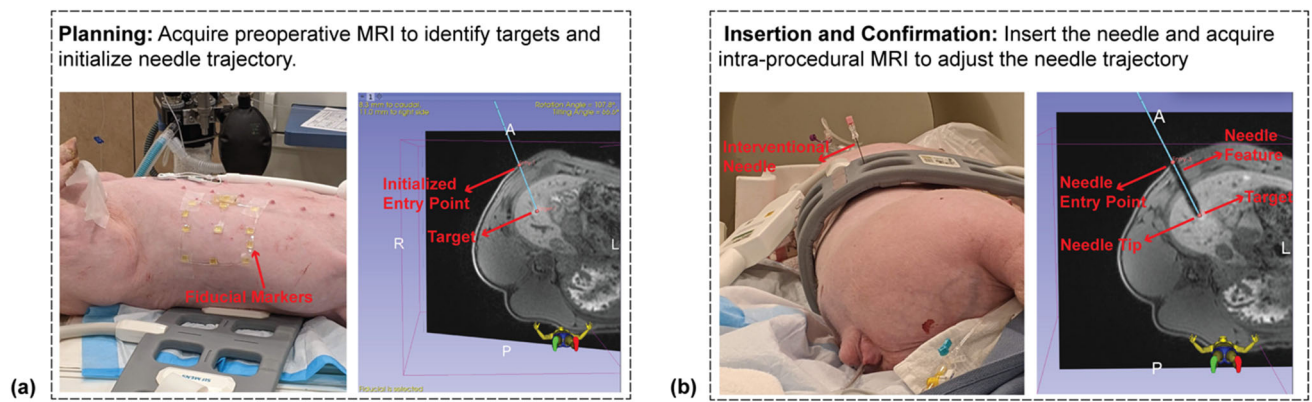


Fig. 1 Manual needle localization workflow for preclinical MRI-guided percutaneous interventions. **a** Planning: Acquire preoperative MR images to localize targets and initialize needle entry point and trajectory. **b** Insertion and Confirmation: Insert the needle and adjust the needle

trajectory based on intermediate confirmation scans until the needle tip reaches the target. Note that needle adjustment/insertion was performed with the subject table moved out of the MRI scanner bore

Table 1 MRI datasets and imaging parameters

	3D T1w Cartesian GRE dataset	2D radial GRE dataset
TR/TE (ms)	3.91/1.23 [OP], 2.46 [IP]	3.8/1.72; 5.08/3; 6.3/2.85
FOV	$237 \times 346 \times 180 \text{ mm}^3$	$300 \times 300 \text{ mm}^2$
Number of slices	120	1
In-plane resolution	$1.35 \times 1.35 \text{ mm}^2$	$1.56 \times 1.56 \text{ mm}^2$
In-plane matrix size	176×256	192×192
Slice thickness	1.5 mm	5 mm
Flip angle	9°	9°
Parallel imaging factor	4	N/A
Acquisition time	13 s (breath held)	100 ms (breathing)
Size of dataset	49 3D volumes for cross-validation (7 3D volumes from each experiment)	490 2D images for cross-validation (70 2D images from each experiment)

TR repetition time, TE echo time, OP out of phase, IP, in phase, FOV field-of-view, N/A not applicable

and supervision of the interventional radiologist, a trained researcher manually annotated the needle feature on the 2D radial GRE images and 3D T1w GRE images to serve as reference segmentation masks. The 3D needle tip and axis references were annotated on the 3D T1w GRE images by marking the 3D coordinates of the needle tip and entry point. The 3D needle tip and axis annotation process were performed twice with a washout period of two weeks in between to assess the human intra-reader variation.

Automatic 3D needle localization pipeline

We proposed a pipeline (Fig. 2) that takes 3D GRE images as input and localizes the needle feature tip and axis in 3D space via a fully automatic process. There were three main steps in the pipeline. Step 1: the 3D Swin UNETR was applied to

the 3D GRE input images and generated the initial 3D needle feature segmentation. The 3D segmentation output was post-processed by a false-positive removal module which calculated the volume of each segmentation object and removed the small ones, as the needle segmentation object had the largest volume compared to the false positives caused by other regions of susceptibility or signal void in the body. Note that false positives connected to the needle segmentation object cannot be removed by this false-positive removal module. Step 2: we performed oblique axial image plane realignment along the main axis of the 3D segmentation output to generate a 2D reformatted image that contains the needle feature. Step 3: the 2D Swin Transformer network was applied to the 2D reformatted image to generate a 2D needle feature segmentation. We localized the 2D needle axis

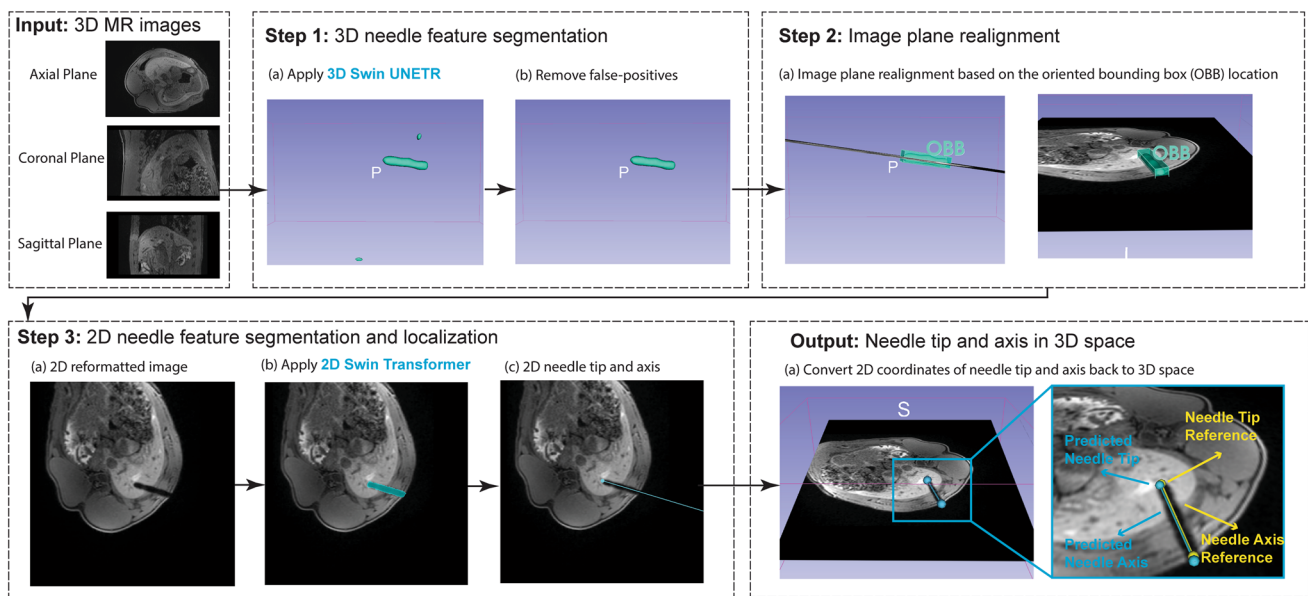


Fig. 2 Diagram of the proposed pipeline. Input: 3D T1w GRE Dixon water images. Step 1: Apply the 3D Swin UNETR for initial 3D needle feature segmentation. Step 2: 2D oblique axial image plane realignment. Step 3: Apply the 2D Swin Transformer on the reformatted 2D

image and localize the needle tip and axis in 2D. Output: Convert the 2D coordinates of the needle tip and axis back to 3D space for 3D visualization

with orthogonal distance regression (ODR) [38]. The intersection of the 2D needle axis and the 2D segmentation mask was identified as the 2D needle feature tip and entry point [20]. We then converted the 2D coordinates of the needle tip and entry point into 3D based on the 2D reformatted image plane position.

To evaluate the necessity of the 2D Swin Transformer network, we compared the performance of the proposed pipeline with the pipeline without the 2D network (step 1 only), which identified the main axis of the 3D segmentation mask as the needle axis, and the intersection of the main axis and the surface of the 3D needle feature segmentation mask as the needle feature tip.

Deep learning networks for needle feature segmentation

We adopted the 3D Swin UNETR [31] (Fig. 3) with pre-trained weights generated by self-supervised learning tasks on publicly available unlabeled CT images of various human body organs without interventional needles [39] and fine-tuned the model using the intra-procedural 3D MR images. We pre-trained the 2D Swin Transformer [40] (Fig. 4) using 2D radial GRE images and then fine-tuned the network using the 2D reformatted images generated by step 2 in the pipeline. Fifteen-fold data augmentation was performed for the training process. To demonstrate the advantages of the 2D and 3D Transformer networks compared with the UNet, we trained

2D UNet and 3D UNet with the same dataset and cross-validation strategy for comparison. The hyperparameters and data augmentation details are shown in Table 2.

Evaluation metrics

To evaluate the needle feature segmentation performance of the 3D Swin UNETR and 2D Swin Transformer, 3D and 2D Dice scores (0–1) of the output segmentations before post-processing were calculated. For 3D needle feature tip and axis localization evaluation, the Euclidean distance between the predicted needle tip and reference needle tip (ε_{tip}) in mm and the angle between the predicted needle axis and needle axis reference (α) in degrees were calculated in 3D space. We performed seven-fold cross-validation using a total of 49 3D volumes (7 from each experiment), where each fold consisted of one experiment's images (7 3D volumes) as the validation set while the training set consisted of images collected from the six remaining experiments (42 3D volumes).

Statistical analysis

We compared differences in the performance (Dice score) of the Swin Transformer-based networks and UNet-based networks, as well as 3D needle localization accuracy (tip and axis error) with and without the 2D Swin Transformer network. For experiments with more than two sets of data

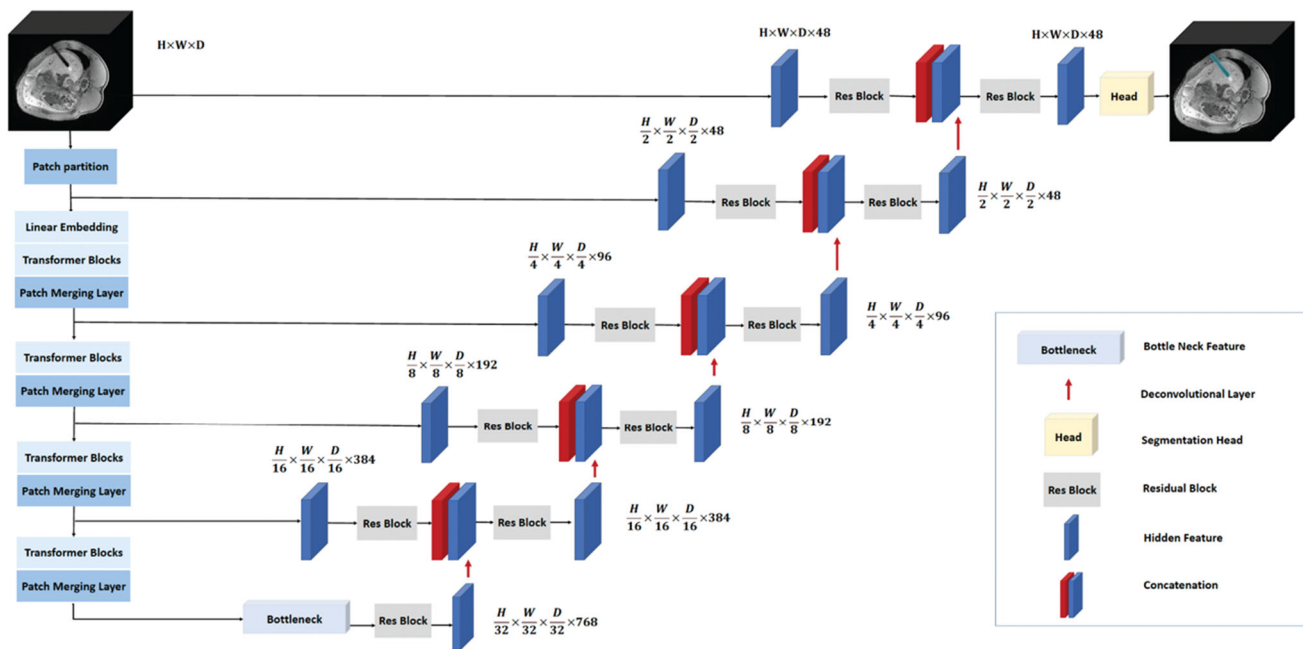


Fig. 3 Overview of the 3D Swin UNETR architecture. The Swin UNETR processed 3D MR images as inputs and generated distinct patches from the input data to establish windows of various sizes for

self-attention calculation. The Swin transformer's encoded feature representations were then transmitted to a CNN decoder through skip connections at various resolutions. W:256, H:256, D:128

samples, the Kruskal–Wallis test was applied first; if the differences were significant across the sets, comparisons were then conducted between pairs of samples using the Wilcoxon signed rank test. Multiple comparisons were accounted for by using Bonferroni correction. A $p < 0.05$ was considered significant.

Results

3D and 2D needle feature segmentation

To assess the benefits of pre-training and data augmentation, we performed an ablation study of different training strategies, and the results are summarized in Supplementary Table S1. The average inference time on one NVIDIA RTX A6000 GPU card (48 GB GPU memory) was 2.14 s per 3D volume for 3D Swin UNETR and 2.67 s per 3D volume for 3D UNet. Representative 3D needle feature segmentation results from 3D Swin UNETR and 3D UNet are shown in Fig. 5. The performance of 3D UNet and 3D Swin UNETR were similar in some cases, while more over-segmentation and under-segmentation were observed in 3D UNet segmentation results. The median [interquartile range (IQR)] of Dice scores was 0.80 [0.11] for 3D Swin UNETR and 0.76 [0.10] for 3D UNet ($p = 0.0073$).

For 2D needle feature segmentation on 2D reformatted images, representative outputs of 2D Swin Transformer and

2D UNet are shown in Fig. 6. The average inference time on the same GPU was 0.011 s per 2D image for the 2D Swin Transformer and 0.016 s per 2D image for the 2D UNet. The median [IQR] of Dice scores was 0.93 [0.04] for 2D Swin Transformer and 0.90 [0.14] for 2D UNet ($p = 0.0110$).

These results (Fig. 7) show statistically significant differences between the performance of 3D Swin UNETR and 3D UNet for 3D needle segmentation, and between the performance of 2D Swin Transformer and 2D UNet for 2D needle segmentation. These results provide evidence that the Swin Transformer-based networks outperform the UNet-based networks in 3D and 2D needle feature segmentation for our application with a limited training dataset.

3D needle localization

The range of needle insertion depth was 1.94–12.26 cm, which is comparable to the skin-to-target length observed in clinical MRI-guided interventions in human subjects (approximately 2–18 cm) [41, 42]. The range of needle insertion angle (angle between the needle and axial plane) was -87.64° to 2.23° . The end-to-end computational time of 3D needle localization was about 6 s per 3D volume for the proposed pipeline and about 4 s for the pipeline without the 2D network. Figure 8 shows example outputs of the pipeline. Volume-rendered displays of the pipeline outputs are shown in Supplementary Video S1.

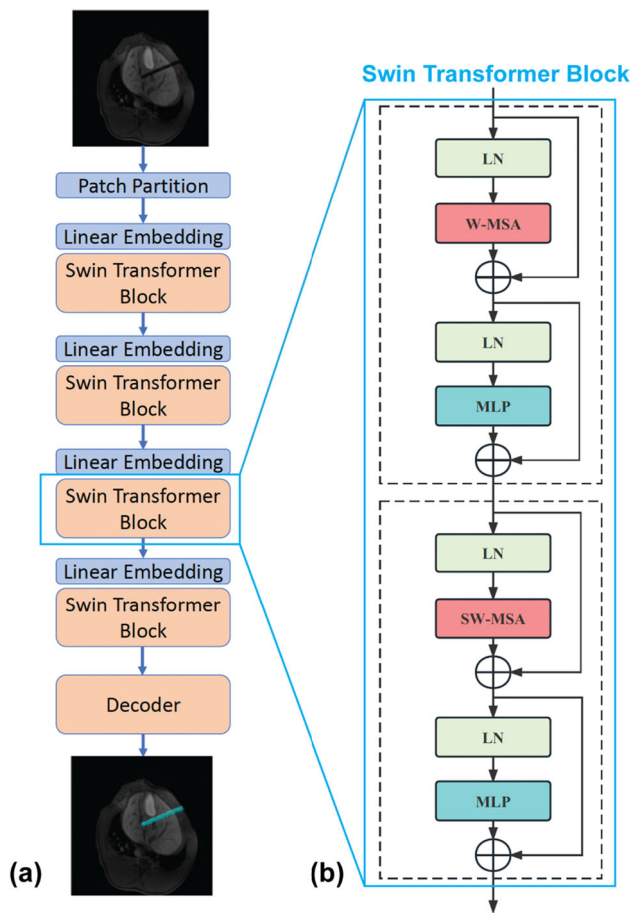


Fig. 4 Overview of the 2D Swin Transformer architecture. **a** The architecture, input 2D MR image, and output 2D segmentation mask of the 2D Swin Transformer. **b** Two successive Swin Transformer Blocks. W-MSA and SW-MSA are multi-head self-attention modules with regular and shifted window configurations, respectively

Figure 9 shows the 3D needle localization results of the proposed pipeline and pipeline without the 2D network (step 1 only) compared with human intra-reader variation as measured by ε_{tip} and α . The ε_{tip} of the proposed pipeline had a median of 1.48 mm (1.09 pixels) and was smaller than the pipeline without the 2D network (median of 1.94 mm; $p = 0.0003$, Wilcoxon signed rank test) and human intra-reader variation (median of 1.70 mm; $p = 0.0085$, Wilcoxon signed rank test). There were no significant differences ($p = 0.5043$, Kruskal–Wallis test) in α between the proposed pipeline (median of 0.98°), the pipeline without the 2D network (median of 0.95°), and human intra-reader variation (median of 1.01°).

Discussion

In this study, we developed a coarse-to-fine automatic deep learning-based pipeline for 3D needle localization on intra-procedural 3D MR images. We used datasets obtained from in vivo MRI-guided interventions in pig livers. The anatomical similarity between pig and human livers is crucial for ensuring that the needle localization pipeline’s development and testing are relevant for future translation to clinical applications in human patients. The proposed pipeline achieved accurate 3D needle localization with a median needle tip localization error of 1.48 mm (1.09 pixels) and a median needle axis localization error of 0.98° . This level of accuracy is sufficient for interventions in the liver (e.g., biopsy or ablation) since clinically relevant lesions typically have a diameter of at least 5–10 mm [41, 43]. With an end-to-end computational time of about 6 s, the proposed pipeline shows the potential to accelerate the current step-and-shoot MRI-guided needle intervention workflow, which involves manual 3D needle localization steps that each take several minutes.

Table 2 Parameters for 3D (3D Swin UNETR and 3D UNet) and 2D neural networks (2D Swin Transformer and 2D UNet)

	3D UNet	3D Swin UNETR	2D UNet	2D Swin Transformer
Loss function	Dice loss		Dice loss	
Learning rate	0.001		Pre-training: 0.001 Fine-tuning: 0.00001	
Optimizer	Adam		Adam	
Batch size	4		8	
Number of epochs	200		Pre-training: 200 Fine-tuning: 100	
Trainable weights	4,808,917	62,186,708	1,625,161	48,962,404
Training time	~ 4 h	~ 5 h	~ 1 h	~ 2 h
Inference time	2.67 s	2.14 s	0.016 s	0.011 s
Data augmentation	15 fold: random rotation, flipping, translation, zooming, additive Gaussian noise			

All networks were implemented using Keras and PyTorch

Fig. 5 Examples of 3D needle feature segmentation outputs before applying the false-positive removal module. 3D needle feature segmentation references (yellow) and neural network predictions of 3D needle feature segmentation (blue) generated by 3D Swin UNETR and 3D UNet are shown. **a** The two networks achieved similar Dice scores. **b** 3D UNet resulted in under-segmentation, while 3D Swin UNETR achieved better performance. **c** 3D UNet resulted in over-segmentation, while 3D Swin UNETR achieved better performance

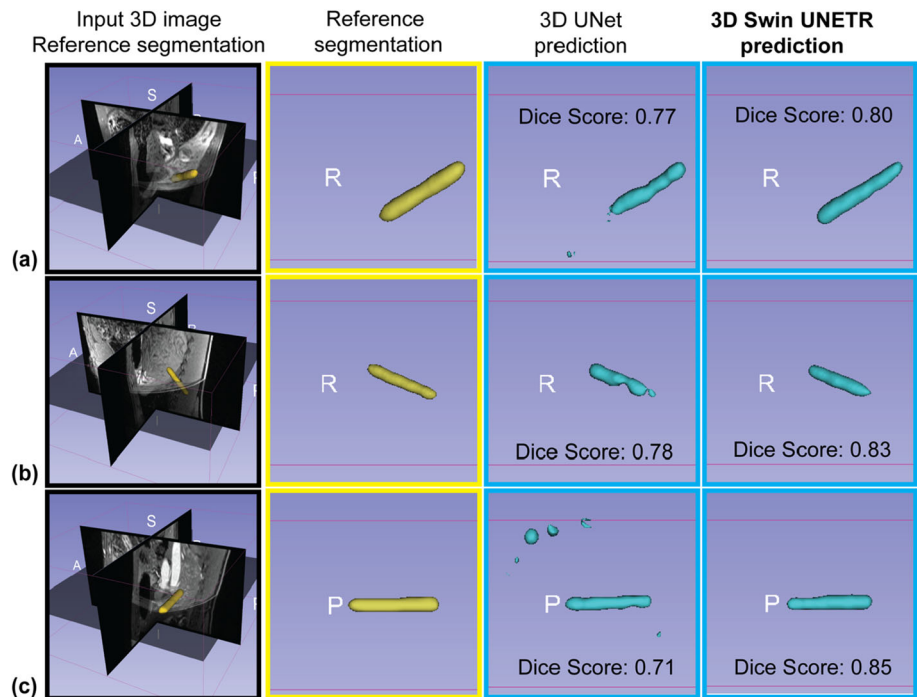
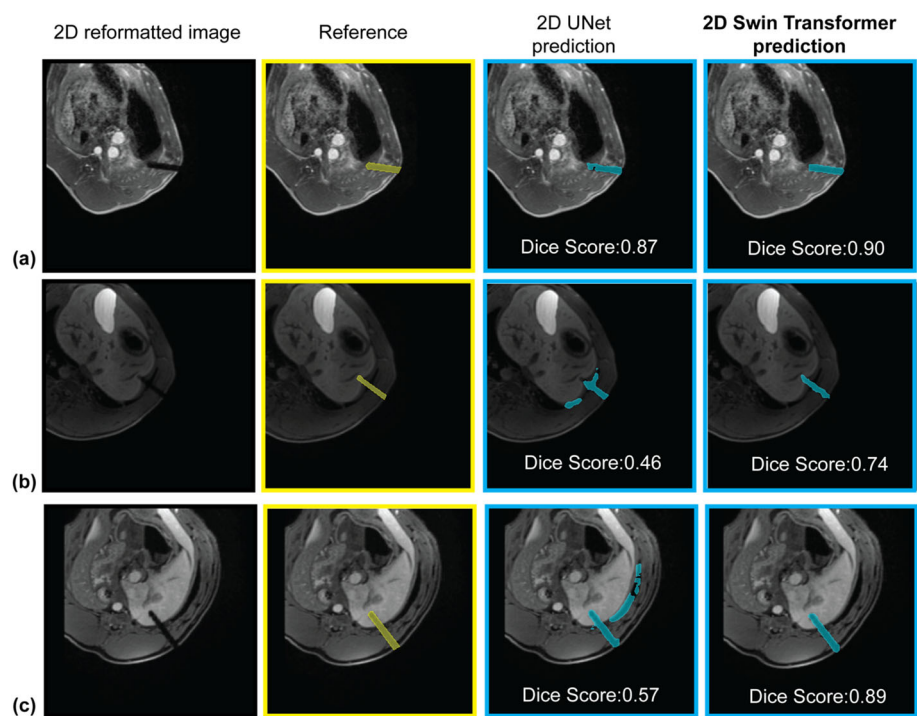


Fig. 6 Examples of 2D needle feature segmentation. The input 2D reformatted image, 2D needle feature segmentation references (yellow), and neural network predictions of 2D needle feature segmentation (blue) generated by 2D Swin Transformer and 2D UNet are shown. **a** The two networks achieved similar Dice scores. **b** 2D UNet resulted in under-segmentation and over-segmentation, while 2D Swin Transformer achieved better performance. **c** 2D UNet resulted in over-segmentation, while 2D Swin Transformer achieved better performance



For 2D and 3D needle feature segmentation, we adopted 2D Swin Transformer and 3D Swin UNETR, respectively. The statistical analyses showed that 3D Swin UNETR and 2D Swin Transformer outperformed the 3D UNet and 2D UNet, which was consistent with the findings of other studies that compared Swin Transformer and UNet-based networks for biomedical image segmentation tasks [39, 44, 45]. These

results demonstrated the advantage of the Swin Transformers in capturing global information when segmenting a small object (i.e., the needle) in a large FOV with complex anatomical structures.

We compared the performance of the proposed pipeline and the pipeline without the 2D network. Under- or over-segmentation of the 3D Swin UNETR still existed due to

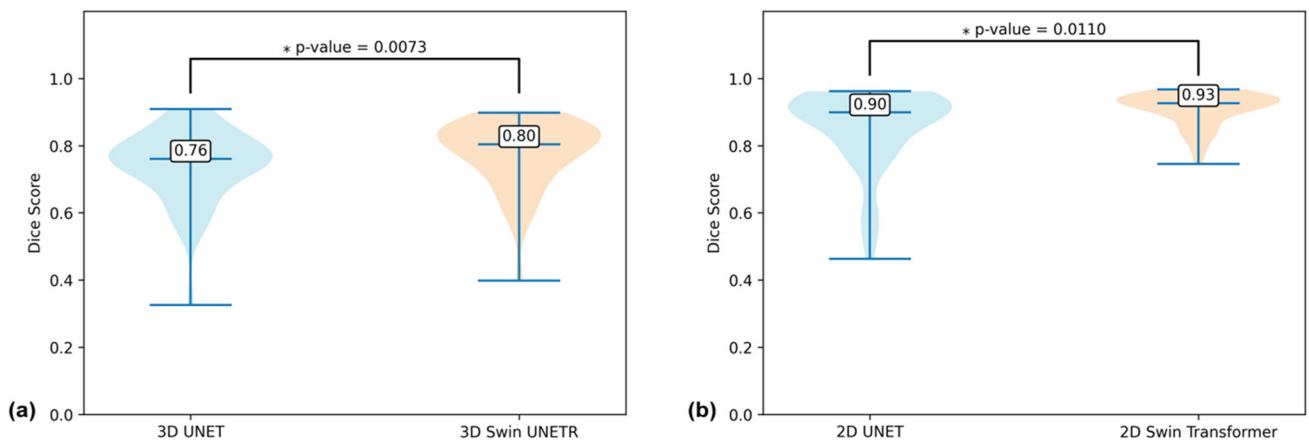


Fig. 7 Needle feature segmentation Dice scores from cross-validation (49 sets of 3D MRI). **a** Violin plots of the Dice scores for 3D needle feature segmentation using 3D UNet and 3D Swin UNETR. **b** Violin plots of the Dice scores for 2D needle feature segmentation using 2D UNet

and 2D Swin Transformer. The numbers shown on the violin plots are the medians of the Dice scores. In the pair-wise comparisons, p -values of the Wilcoxon signed rank test are shown on the connecting lines. * indicates $p < 0.05$

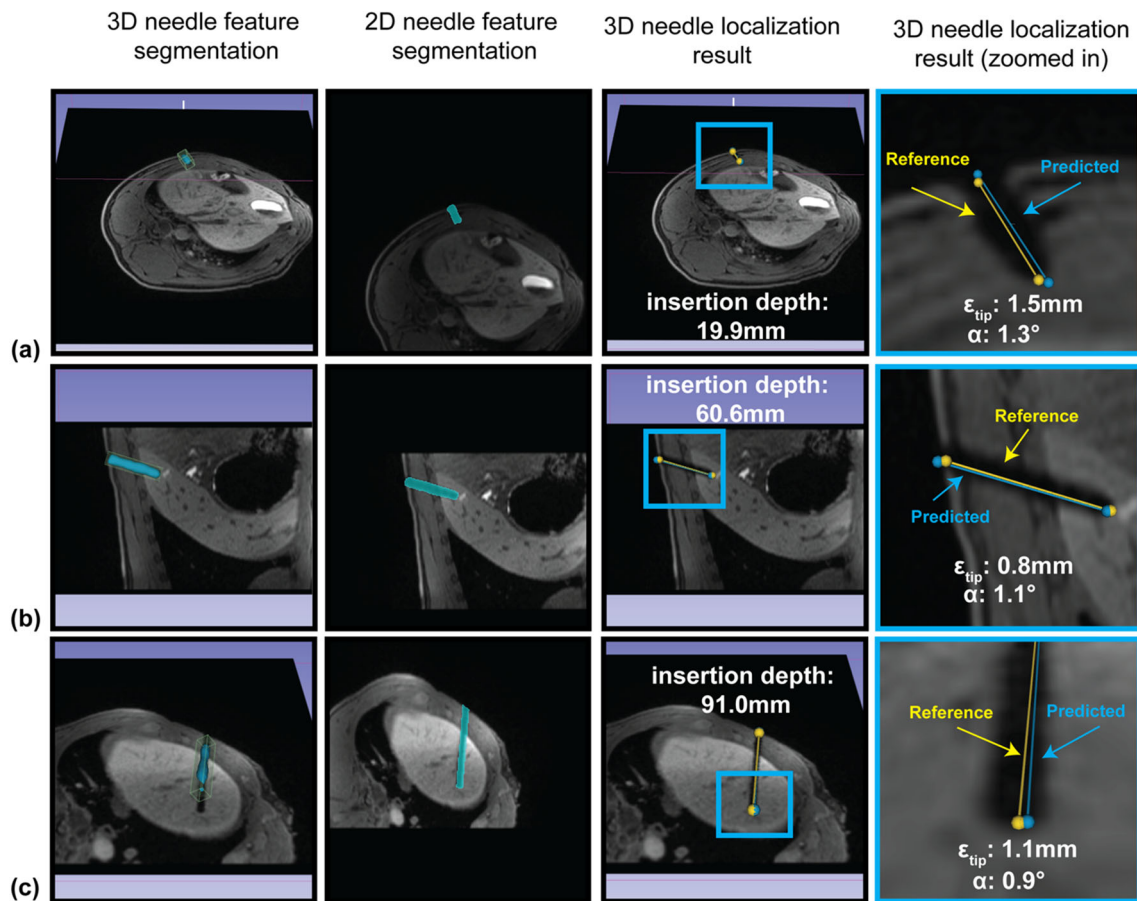


Fig. 8 Example outputs from the proposed 3D needle localization pipeline. **a** Shallow insertion depth around 20 mm. **b** Moderate insertion depth around 60 mm. **c** Deeper insertion depth around 90 mm. 3D needle feature segmentation: 3D needle feature segmentation shown with the 2D reformatted image plane in 3D Slicer. 2D needle feature

segmentation: 2D needle feature segmentation shown on the 2D reformatted image. 3D needle localization results: Predicted (blue) and reference (yellow) needle tip and axis in 3D space. The needle tip error (ϵ_{tip} ; mm) and needle axis error (α ; deg) are reported for each example

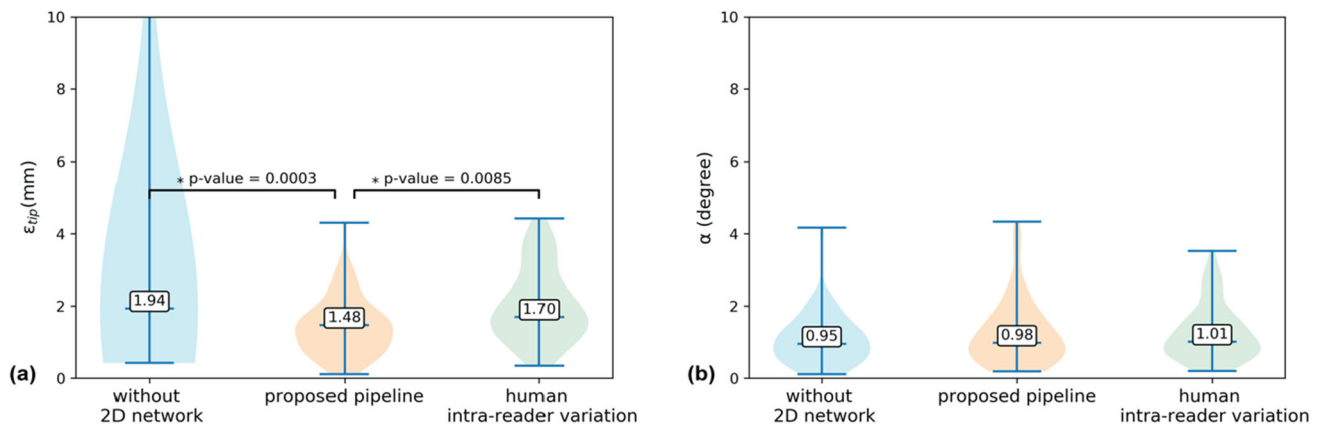


Fig. 9 Automatic 3D needle localization results from cross-validation (49 sets of 3D MRI). **a** Violin plots of needle tip localization error (ϵ_{tip}) and **b** violin plots of needle axis localization error (α) of the proposed pipeline, pipeline without 2D network, and human intra-reader

variation. The numbers shown on the violin plots are the medians of the results. In the pair-wise comparisons, p-values of the Wilcoxon signed rank test are shown on the connecting lines. * indicates $p < 0.05$

the limitation of the size of the 3D MRI training dataset. The under- or over-segmentation usually appeared near the needle tip and entry points and therefore had little effect on the needle axis localization but could lead to large needle tip localization errors in the pipeline without a 2D network. Therefore, combining the 2D network in the pipeline was necessary to compensate for the under- or over-segmentation result of the 3D Swin UNETR. In the future, the 2D network might become unnecessary if the 3D network achieves the required accuracy for guiding interventions with additional training data.

There were limitations to this study. Firstly, due to the limited size of the intra-procedural 3D MRI dataset, the training of the networks was affected, and all the results reported here were from cross-validation experiments. In the future, more interventional experiments will be conducted to acquire more data. The additional data will expand the training dataset and enable independent testing for a more comprehensive assessment of the pipeline's performance. Secondly, the reference of the needle tip and axis was annotated by one observer with a washout period of two weeks to assess the human intra-reader variation. Future work can consider multiple observers and use majority voting for needle tip localization reference creation. Thirdly, inline deployment and prospective demonstration of the proposed pipeline in the context of a procedure was not yet achieved. Future work will focus on integrating and testing the proposed pipeline in in vivo MRI-guided interventions.

Conclusion

In this work, we developed a deep learning-based pipeline for automatic 3D needle localization on intra-procedural 3D MR

images. The pipeline had a coarse-to-fine structure where it adopted 3D Swin UNETR for initial segmentation of the 3D needle feature and 2D Swin Transformer for fine segmentation of the needle feature in the 2D reformatted image plane. The proposed pipeline achieved rapid and accurate 3D needle localization within the range of expert human performance and thus has potential to improve MRI-guided percutaneous interventions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11548-024-03077-3>.

Acknowledgements This work received funding support from Siemens Medical Solutions USA, the Department of Radiological Sciences at UCLA, and the National Institute of Biomedical Imaging and Bio-engineering of the National Institutes of Health under award number R01EB031934. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declarations

Conflict of interest This study was supported in part by Siemens Medical Solutions USA, Inc. in the form of a research grant (X.L., D.S.L., H.H.W.). Siemens had no role in the study design, data collection, analysis, or preparation of this manuscript.

Ethical Approval This article does not contain any studies with human participants. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution at which the studies were conducted.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cleary K, Peters TM (2010) Image-guided interventions: technology review and clinical applications. *Annu Rev Biomed Eng* 12:119–142. <https://doi.org/10.1146/annurev-bioeng-070909-105249>
2. Mirota DJ, Ishii M, Hager GD (2011) Vision-based navigation in image-guided interventions. *Annu Rev Biomed Eng* 13:297–319. <https://doi.org/10.1146/annurev-bioeng-071910-124757>
3. Neshat H, Cool DW, Barker K, Gardi L, Kakani N, Fenster A (2013) A 3D ultrasound scanning system for image guided liver interventions. *Med Phys* 40:112903. <https://doi.org/10.1118/1.4824326>
4. Spinczyk D (2015) Towards the clinical integration of an image-guided navigation system for percutaneous liver tumor ablation using freehand 2D ultrasound images. *Comput Aided Surg* 20:61–72. <https://doi.org/10.3109/10929088.2015.1076043>
5. Yu NC, Chaudhari V, Raman SS, Lassman C, Tong MJ, Busuttill RW, Lu DSK (2011) CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis. *Clin Gastroenterol Hepatol* 9:161–167. <https://doi.org/10.1016/j.cgh.2010.09.017>
6. Stattaus J, Kuehl H, Ladd S, Schroeder T, Antoch G, Baba HA, Barkhausen J, Forsting M (2007) CT-guided biopsy of small liver lesions: visibility, artifacts, and corresponding diagnostic accuracy. *Cardiovasc Intervent Radiol* 30:928–935. <https://doi.org/10.1007/s00270-007-9023-8>
7. Campbell-Washburn AE, Tavallaei MA, Pop M, Grant EK, Chubb H, Rhode K, Wright GA (2017) Real-time MRI guidance of cardiac interventions. *J Magn Reson Imaging* 46:935–950. <https://doi.org/10.1002/jmri.25749>
8. Kaye EA, Granlund KL, Morris EA, Maybody M, Solomon SB (2015) Closed-bore interventional MRI: percutaneous biopsies and ablations. *Am J Roentgenol* 205:W400–W410. <https://doi.org/10.2214/AJR.15.14732>
9. Rempp H, Clasen S, Pereira PL (2012) Image-Based monitoring of magnetic resonance-guided thermoablative therapies for liver tumors. *Cardiovasc Intervent Radiol* 35:1281–1294. <https://doi.org/10.1007/s00270-011-0227-6>
10. Weiss CR, Nour SG, Lewin JS (2008) MR-guided biopsy: a review of current techniques and applications. *J Magn Reson Imag* 27:311–325. <https://doi.org/10.1002/jmri.21270>
11. Kurumi Y, Tani T, Naka S, Shiomi H, Shimizu T, Abe H, Endo Y, Morikawa S (2007) MR-guided microwave ablation for malignancies. *Int J Clin Oncol* 12:85–93. <https://doi.org/10.1007/s10147-006-0653-7>
12. Raj SD, Agrons MM, Woodtichartpreecha P, Kalambo MJ, Dogan BE, Le-Petross H, Whitman GJ (2019) MRI-guided needle localization: Indications, tips, tricks, and review of the literature. *Breast J* 25:479–483. <https://doi.org/10.1111/tbj.13246>
13. DeAngelis GA, Moran RE, Fajardo LL, Mugler JP, Christopher JM, Harvey JA (2000) MRI-guided needle localization: technique. *Semin Ultrasound, CT MRI* 21:337–350. [https://doi.org/10.1016/S0887-2171\(00\)90028-3](https://doi.org/10.1016/S0887-2171(00)90028-3)
14. Development and validation of a real-time reduced field of view imaging driven by automated needle detection for MRI-guided interventions. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7625/762515/Development-and-validation-of-a-real-time-reduced-field-of-view-imaging-driven-by-automated-needle-detection-for-MRI-guided-interventions>
15. Daanen V, Coste E, Sergent G, Godart F, Vasseur C, Rousseau J (2000) Accurate localization of needle entry point in interventional MRI. *J Magn Reson Imaging* 12:645–649. [https://doi.org/10.1002/1522-2586\(200010\)12:4%3c645::aid-jmri19%3e3.0.co;2-3](https://doi.org/10.1002/1522-2586(200010)12:4%3c645::aid-jmri19%3e3.0.co;2-3)
16. Meinhold W, Martinez DE, Oshinski J, Hu A-P, Ueda J (2020) A direct drive parallel plane piezoelectric needle positioning robot for MRI guided intraspinal injection. *IEEE Trans Biomed Eng* 68:807–814
17. Morris EA, Liberman L, Dershaw DD, Kaplan JB, LaTrenta LR, Abramson AF, Ballon DJ (2002) Preoperative MR imaging—guided needle localization of breast lesions. *Am J Roentgenol* 178:1211–1220. <https://doi.org/10.2214/ajr.178.5.1781211>
18. Franco E, Ristic M, Rea M, Gedroyc WMW (2016) Robot-assistant for MRI-guided liver ablation: a pilot study. *Med Phys* 43:5347–5356. <https://doi.org/10.1118/1.4961986>
19. Wu D, Li G, Patel N, Yan J, Monfaredi R, Cleary K, Iordachita I (2019) Remotely actuated needle driving device for MRI-guided percutaneous interventions. In: 2019 International Symposium on Medical Robotics (ISMR), pp 1–7
20. Li X, Young AS, Raman SS, Lu DS, Lee Y-H, Tsao T-C, Wu HH (2020) Automatic needle tracking using mask R-CNN for MRI-guided percutaneous interventions. *Int J CARS* 15:1673–1684. <https://doi.org/10.1007/s11548-020-02226-8>
21. Li X, Lee Y-H, Lu DS, Tsao T-C, Wu HH (2021) Physics-driven mask R-CNN for physical needle localization in MRI-guided percutaneous interventions. *IEEE Access* 9:161055–161068. <https://doi.org/10.1109/ACCESS.2021.3128163>
22. Weine J, Rothgang E, Wacker F, Weiss CR, Maier F (2018) Passive needle tracking with deep convolutional neural nets for MR-guided percutaneous interventions. In: Proceedings of 12th Interventional MRI Symposium Oct. p 53
23. Lee E-J, Farzinfard S, Yarmolenko P, Cleary K, Monfaredi R (2023) Toward robust partial-image based template matching techniques for MRI-guided interventions. *J Digit Imaging* 36:153–163. <https://doi.org/10.1007/s10278-022-00716-6>
24. Mehrtash A, Ghafoorian M, Pernelle G, Ziaei A, Heslinga FG, Tuncali K, Fedorov A, Kikinis R, Tempny CM, Wells WM, Abolmaesumi P, Kapur T (2019) Automatic needle segmentation and localization in MRI with 3-D convolutional neural networks: application to MRI-targeted prostate biopsy. *IEEE Trans Med Imag* 38:1026–1036. <https://doi.org/10.1109/TMI.2018.2876796>
25. Anh LQ, Ha LM, Van Walsum T, Moelker A, Hang DV, Phuong PC, Thanh VD (2022) Needle localization and segmentation for radiofrequency ablation of liver tumors under CT image guidance. In: 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp 2015–2021
26. Yang H, Shan C, Bouwman A, Kolen AF, de With PHN (2021) Efficient and robust instrument segmentation in 3D ultrasound using patch-of-interest-fusenet with hybrid loss. *Med Image Anal* 67:101842. <https://doi.org/10.1016/j.media.2020.101842>
27. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK (2022) Medical image segmentation using deep learning: a survey. *IET Image Proc* 16:1243–1267. <https://doi.org/10.1049/ipr2.12419>
28. Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging* 32:582–596. <https://doi.org/10.1007/s10278-019-00227-x>
29. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houshy N (2021) An image is worth 16x16 words: transformers for image recognition at scale

30. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z, Zhang Y, Tao D (2023) A survey on vision transformer. *IEEE Trans Pattern Anal Mach Intell* 45:87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>
31. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D (2022) Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi A, Bakas S (eds) *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries*. Springer International Publishing, Cham, pp 272–284
32. Worakitsitatorn A, Lu DS, Lee MW, Asvadi NH, Moshksar A, Yuen AD, McWilliams J, Raman SS (2020) Percutaneous thermal ablation of subcapsular hepatocellular carcinomas: influence of tumor-surface contact and protrusion on therapeutic efficacy and safety. *Eur Radiol* 30:1813–1821. <https://doi.org/10.1007/s00330-019-06497-x>
33. Lee JK, Siripongsakun S, Bahrami S, Raman SS, Sayre J, Lu DS (2016) Microwave ablation of liver tumors: degree of tissue contraction as compared to RF ablation. *Abdom Radiol (NY)* 41:659–666. <https://doi.org/10.1007/s00261-016-0725-8>
34. Yuan F, Wei SH, Konecny GE, Memarzadeh S, Suh RD, Sayre J, Lu DS, Raman SS (2021) Image-guided percutaneous thermal ablation of oligometastatic ovarian and non-ovarian gynecologic tumors. *J Vasc Interv Radiol* 32:729–738. <https://doi.org/10.1016/j.jvir.2021.01.270>
35. Tan N, Lin W-C, Khoshnoodi P, Asvadi NH, Yoshida J, Margolis DJA, Lu DSK, Wu H, Sung KH, Lu DY, Huang J, Raman SS (2017) In-bore 3-T MR-guided transrectal targeted prostate biopsy: prostate imaging reporting and data system version 2-based diagnostic performance for detection of prostate cancer. *Radiology* 283:130–139. <https://doi.org/10.1148/radiol.2016152827>
36. Felker ER, Lee-Felker SA, Feller J, Margolis DJ, Lu DS, Princenthal R, May S, Cohen M, Huang J, Yoshida J, Greenwood B, Kim HJ, Raman SS (2016) In-bore magnetic resonance-guided transrectal biopsy for the detection of clinically significant prostate cancer. *Abdom Radiol (NY)* 41:954–962. <https://doi.org/10.1007/s00261-016-0750-7>
37. Pieper S, Halle M, Kikinis R (2004) 3D slicer. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat no. 04EX821). pp 632–635 Vol. 1
38. Poch J, Villaescusa I (2012) Orthogonal distance regression: a good alternative to least squares for modeling sorption data. *J Chem Eng Data* 57:490–499. <https://doi.org/10.1021/je201070u>
39. Tang Y, Yang D, Li W, Roth HR, Landman B, Xu D, Nath V, Hatamizadeh A (2022) Self-supervised pre-training of swin transformers for 3d medical image analysis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp 20730–20740
40. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Montreal, QC, Canada, pp 9992–10002
41. Moche M, Heinig S, Garnov N, Fuchs J, Petersen T-O, Seider D, Brandmaier P, Kahn T, Busse H (2016) Navigated MRI-guided liver biopsies in a closed-bore scanner: experience in 52 patients. *Eur Radiol* 26:2462–2470. <https://doi.org/10.1007/s00330-015-4097-1>
42. Hoffmann R, Thomas C, Rempp H, Schmidt D, Pereira PL, Claussen CD, Clasen S (2012) Performing MR-guided biopsies in clinical routine: factors that influence accuracy and procedure time. *Eur Radiol* 22:663–671. <https://doi.org/10.1007/s00330-011-2297-x>
43. Parekh PJ, Majithia R, Diehl DL, Baron TH (2015) Endoscopic ultrasound-guided liver biopsy. *Endosc Ultrasound* 4:85–91. <https://doi.org/10.4103/2303-9027.156711>
44. Gomes R, Pham T, He N, Kamrowski C, Wildenberg J (2023) Analysis of Swin-UNet vision transformer for Inferior Vena Cava filter segmentation from CT scans. *Artif Intell Life Sci* 4:100084. <https://doi.org/10.1016/j.aillsci.2023.100084>
45. Wu J, Fu R, Fang H, Zhang Y, Xu Y (2023) Medsegdiff-v2: diffusion based medical image segmentation with transformer. arXiv preprint arXiv:230111798

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.