This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

Learning-based Hybrid Needle Tip Tracking and Visualization Framework for Robotic Ultrasound-guided Interventions

Junling Fu†, *Member, IEEE*, Zijian Cai†, Runing Xiao, Giancarlo Ferrigno, *Senior Member, IEEE*, Alberto Redaelli, and Elena De Momi, *Senior Member, IEEE*

***Abstract*—Ultrasound (US)-guided needle-based interventions have been extensively employed in clinical practice for biopsies, regional anesthesia, and tumor ablation, considering the advantages of real-time imaging, portability, and non-ionizing radiation. Precise localization and tracking of the needle tip are critical for ensuring the successful needle placement while minimizing the risk of injury to surrounding anatomical structures. However, achieving real-time and reliable needle tip tracking remains challenging due to the inherent properties of US images, such as low spatial resolution, speckle noise, and imaging artifacts that obscure the needle tip. This study proposes a hybrid learning-based framework for accurately and robustly localizing the needle tip within the US image plane during robot-assisted interventional procedures. The framework comprises a dual-branch template matching-based detection module, a Convolutional Neural Network (CNN)-based detection module, and a "dynamic template selection and updating" mechanism to enhance overall tracking performance. Additionally, to recover the needle tip invisibility caused by unanticipated movements, an active exploration strategy based on a Gaussian Process (GP) model and Bayesian Optimization (BO) is proposed for efficient and accurate robotic US probe pose adjustment. The effectiveness of the proposed method for needle tip tracking performance has been evaluated through a series of experiments conducted on *ex vivo* porcine liver samples under varying insertion angles and velocities. Experimental results demonstrate that the proposed framework achieves a maximum median localization error of 1.17 mm and an Interquartile Range (IQR) of 1.28 mm across all test conditions. Furthermore, in scenarios requiring needle tip visibility recovery, the proposed BO-based strategy achieved a median localization error of 1.52 mm and an IQR of 0.95 mm.**

***Index Terms*—Medical Robot, Ultrasound Imaging, Needle Tip Detection, Robot Ultrasound, Optimization Control.**

I. INTRODUCTION

Needle-based interventions are commonly used in applications such as biopsy, ablation, regional anesthesia, and intervention radiology [1], [2]. Precise needle placement is critical to ensure the target position is reached while avoiding damaging sensitive anatomical structures like major vessels, nerves, and ribs, directly impacting subsequent diagnostic and therapeutic outcomes [3]. To obtain the needle position during the intervention procedures, multiple imaging modalities have been utilized, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and fluoroscopy [4]. Among these modalities, MRI offers high-resolution imaging. However, it typically requires custom, non-ferromagnetic needles, which are often cost-prohibitive. Additionally, both CT and fluoroscopy expose patients to radiation, and CT is not suitable for real-time imaging during intraoperative procedures. In contrast, US imaging is a widely used modality for guiding needle-based interventions due to its real-time capability, radiation-free, and cost-effectiveness merits [2], [3], [5].

Conventional US-guided interventions require clinicians to simultaneously manipulate both the US probe and the needle. The accuracy of needle placement heavily relies on the operator's ability to interpret raw US images, which are often characterized by low contrast, poor image quality, and susceptibility to speckle noise, artifacts, and reverberation. These limitations make real-time visual identification of the needle's position challenging. To address this, various advanced techniques have been developed for accurate needle tip tracking, which can be categorized into hardware-based and imaging-based approaches [2], [6]. Typically, hardware-based methods employ external tracking systems or embedded sensors, such as optical [7], [8], electromagnetic (EM) [9], or magnetic modalities [10] to localize the needle tip within the US image plane. While these methods offer high accuracy and potential for miniaturization, some limitations are present, including interference from the EM field generators with other medical devices, occlusion in optical systems, and increased complexity due to calibration requirements.

In contrast, imaging-based techniques are typically more adaptable, easier to integrate with existing systems, and do not require additional hardware, making them a more flexible solution in clinical practice [11]. Image processing-based techniques have been explored to distinguish the needle from the US image and accurately localize the needle tip, such as the Hough transform, filtering, and template matching. In [12], a Gabor filter combined with thresholding and morphological operations was designed to extract fine needle features. Specifically, the Random Sample Consensus (RANSAC) algorithm was then utilized to estimate the needle insertion angle,

---

This work was funded by the project "THE - Tuscany Health Ecosystem" (ECS00000017, CUP I53C22000780001), specifically through Spoke 9, "Robotics and Automation for Health" for the "REDIT - Robot-assisted Remote Echography for Diagnosis and Treatment". Also, it was partially funded by the Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN)-Bando 2022, under the project "FutuRO" (Grant Number: 2022342 MEF).

Junling Fu, Runing Xiao, Giancarlo Ferrigno, Alberto Redaelli, and Elena De Momi are with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133, Milan, Italy.

Zijian Cai is with the Department of Mechanical Engineering, Politecnico di Milano, 20133, Milan, Italy.

†, These authors contributed equally to this work. Corresponding author: Junling Fu, junling.fu@polimi.it

from which the needle tip position was subsequently derived. Similarly, a second-order derivative of the Gaussian filter has been used to suppress noise in raw US images and enhance the visibility of the biopsy needle in transrectal ultrasound [13]. To improve the robustness of needle tip tracking, a statistical filtering-based detection approach was proposed in [14], incorporating online Region of Interest (ROI) adaptation for needle axis estimation, alongside a particle filter to track the needle tip by mitigating noise and outliers. Typically, these methods generally estimate the needle tip based on the detection or segmentation of the needle shaft. However, the needle shaft is not always clearly visible, and the parameters of these filters often require careful tuning. Template matching techniques have also been implemented, leveraging raw pixel values to measure the similarity between a target template and candidate regions for needle tip localization [15]. Nonetheless, significant variations of the needle tip's appearance within biological tissue can substantially degrade detection performance and limit adaptability in complex and noisy scenarios.

Compared to traditional image processing techniques, deep neural networks are capable of extracting rich, multi-level features that significantly enhance the accuracy of needle axis segmentation and tip localization in US images. In [16], a unified architecture combining a Fully Convolutional Network (FCN) with a fast Region-based Convolutional Neural Network (R-CNN) was proposed to estimate the needle trajectory based on intensity-invariant features. Similarly, [17] introduced a novel CNN framework featuring a dual-channel encoder and a single-channel decoder for needle segmentation, which exploits motion cues across adjacent US frames. Despite these advancements, segmentation-based approaches still struggle to identify the full needle axis when it is only partially visible, which ultimately results in tracking failure.

Additionally, Deep Learning-based methods for direct surgical instrument tip detection in US imaging have been explored [18], [19]. For example, to deal with axis invisible and low-intensity issues of the needle, a hybrid deep neural network architecture was proposed by combining a CNN module for detecting subtle intensity variations with a long short-term memory (LSTM) recurrent module for needle tip location prediction [20]. However, the LSTM component is sensitive to the quality and consistency of input sequences, limiting its robustness in noisy and complex environments. To further enhance detection performance, more advanced Deep Learning-based architectures have been introduced to improve needle feature representation. Yan et al. [21] proposed a visual tracking framework based on a Transformer architecture, designed to reinforce and propagate needle tip appearance features. Gaussian-shaped masks were applied to emphasize target regions while suppressing background noise, and the model employed a discriminative correlation filter (DCF) along with a computationally expensive bounding box regression head for final localization. Although this method demonstrates significantly improved performance in offline validation, as discussed in the conclusion, its reliance on consistent needle tip appearance and high computational cost hinders its effectiveness under rapid appearance changes and precludes its use in real-time tracking scenarios.

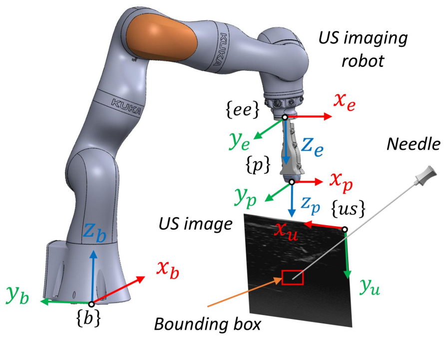

Fig. 1. System components and coordinate transformations for Robotic US-guided needle tip detection. $\{\boldsymbol{us}\}$, $\{\boldsymbol{p}\}$, $\{\boldsymbol{b}\}$, and $\{\boldsymbol{ee}\}$ represent the US image, probe, robot base, and robot end-effector frames, respectively.

To enhance the accuracy and visibility of the needle tip, this study proposes a comprehensive framework that integrates a learning-based hybrid needle tip tracking architecture with robotic systems for US-guided interventions. The proposed framework is designed to ensure robust and precise needle tip tracking and localization, accounting for common challenges such as image noise, artifacts, and bright speckle patterns. Additionally, an active robotic exploration strategy is incorporated to effectively reorient the needle tip within the imaging plane of the US probe due to unforeseen conditions. The main contributions of this study can be summarized as follows:

- A template-matching-based module is implemented for efficient needle tip detection from the US images, utilizing a deep network-based feature extractor and a dual-branch Siamese Region Proposal Network architecture.
- A hybrid tracking framework is developed by combining the proposed template-matching and CNN-based detection modules. Specifically, the bounding box selection and template updating strategies have been proposed to achieve efficient needle tip tracking performance.
- A Bayesian Optimization-based strategy is proposed for the US probe pose adjustment to facilitate needle tip re-localization and restore the visibility within the US image.

The remainder of this work is organized as follows. Section II illustrates the details of the proposed learning-based hybrid needle tip tracking and visualization framework. Then, Section III presents the system design, dataset preparation, training, and the experimental protocol. Section IV reports the experiment results. Finally, Section V concludes this work.

II. METHODOLOGY

A. Coordinates Transformation and Robot Controller

1) Robotic US System and Coordinates Transformation: To perform precise needle tip detection from US images during interventions, calibration between the various frames should be implemented, as depicted in Fig. 1. Specifically, the detected needle tip's position, P'^{b}_{tip}, within the US image in the $\{\boldsymbol{b}\}$ frame is calculated as:

$$P'^{b}_{tip} = {}^{b}_{us}\mathbf{T}P^{us}_{tip} \tag{1}$$

where ${}^{b}_{us}\mathbf{T} \in \mathbb{R}^{4\times4}$ represents the the transformation matrix between the frames $\{\boldsymbol{us}\}$ and $\{\boldsymbol{b}\}$. $P^{us}_{tip} \in \mathbb{R}^{4}$ denotes the pixel

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

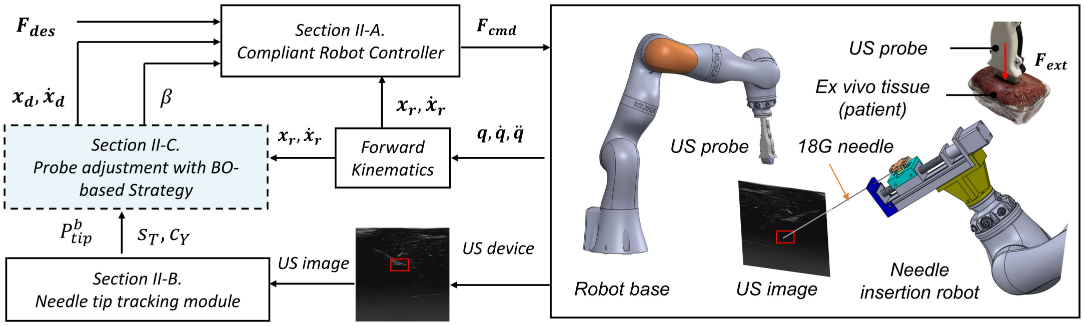

Fig. 2. Details of the proposed control framework for needle tip tracking and visualization during robotic US-guided needle interventions.

position of the detected needle tip in $\{\boldsymbol{us}\}$ frame, which is the center point of the bounding box in the image. Moreover, the matrix ${}^{b}_{us}\mathbf{T}$ can be calculated according to the relationship of the frames, which is expressed as below:

$$ {}^{b}_{us}\mathbf{T} = {}^{b}_{ee}\mathbf{T}{}^{ee}_{p}\mathbf{T}{}^{p}_{us}\mathbf{T} \tag{2}$$

where ${}^{b}_{ee}\mathbf{T} \in \mathbb{R}^{4\times4}$ is calculated using the forward kinematics from $\{\boldsymbol{b}\}$ to $\{\boldsymbol{ee}\}$. ${}^{ee}_{p}\mathbf{T} \in \mathbb{R}^{4\times4}$ denotes the transformation matrix from the end effector frame $\{\boldsymbol{ee}\}$ to the center of the US probe $\{\boldsymbol{p}\}$, which can be derived utilizing the "4-points-based calibration" procedure provided by the robot with the teaching pendant. Furthermore, ${}^{p}_{us}\mathbf{T}$ is determined by the geometrical relationship and parameters of the US imaging system [22], which can be calculated as follows:

$$ {}^{p}_{us}\mathbf{T} = \begin{bmatrix} -\frac{L_p}{W_{us}} & 0 & 0 & \frac{L_p}{2} \\ 0 & 0 & -1 & 0 \\ 0 & \frac{D_p}{H_{us}} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where L_p and D_p are the footprint length, i.e., the sensor elements on the linear US probe (the linear US probe length in this work is 40 mm) and the B-model imaging depth in Cartesian space (can be specified on the US machine interface), respectively. W_{us} and H_{us} are the pixel width and height of the US image, respectively. Moreover, the origin point of the US image is located at the upper-left point; hence, a translation of $\frac{L_p}{2}$ is applied for the coordinate transformation from the upper left point to the center of the US probe.

2) Compliant Control for Robotic US Imaging: To achieve safe contact between the robot and anatomical structures, a compliant controller, the Cartesian impedance controller, has been utilized [23]. Specifically, the external contact force, $\boldsymbol{F_{ext}}$, between the robot and the environment can be expressed by a *Mass-Spring-Damper* (*MSD*) system:

$$ \boldsymbol{M}_c\ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}_c\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}_c\tilde{\boldsymbol{x}} = \boldsymbol{F_{ext}} \tag{4}$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{x_d} - \boldsymbol{x_r}, \tilde{\boldsymbol{x}} \in \mathbb{R}^m$ represents the error between the desired position $\boldsymbol{x_d}$ and the measured robot position $\boldsymbol{x_r}$. $\dot{\tilde{\boldsymbol{x}}}$ and $\ddot{\tilde{\boldsymbol{x}}} \in \mathbb{R}^m$ are velocity and acceleration vectors. $\boldsymbol{M_c}, \boldsymbol{D_c}, \boldsymbol{K_c} \in \mathbb{R}^{m\times m}$ are the positive definite mass, stiffness, and damping parameters, respectively. Furthermore, a proper desired contact force, $\boldsymbol{F_{des}}$, needs to be set for stable and clear US image acquisition. Ultimately, considering the dynamics model of an n-degree-of-freedoms (DoFs) serial robot in the m-DoFs Cartesian space (gravity force has been compensated), the following formula for the robot control can be obtained:

$$ \boldsymbol{M_r}(\boldsymbol{x_r})\ddot{\boldsymbol{x}}_{\boldsymbol{r}} + \boldsymbol{C_r}(\boldsymbol{x_r}, \dot{\boldsymbol{x}}_{\boldsymbol{r}})\dot{\boldsymbol{x}}_{\boldsymbol{r}} + \boldsymbol{F_{ext}} + \boldsymbol{F_{des}} = \boldsymbol{F_{cmd}} \tag{5}$$

where $\boldsymbol{M_r}$ and $\boldsymbol{C_r} \in \mathbb{R}^{m\times m}$ are the symmetric and positive definite mass matrix, the Coriolis and Centrifugal matrix of the robot, respectively. $\boldsymbol{x_r}, \dot{\boldsymbol{x}}_{\boldsymbol{r}}, \ddot{\boldsymbol{x}}_{\boldsymbol{r}} \in \mathbb{R}^m$ are the measured position, velocity, and acceleration terms of the robot in Cartesian space; $\boldsymbol{F_{ext}} \in \mathbb{R}^m$ is the external contact force to ensure high-quality US image acquisition (to ensure the US probe is in contact with the phantom or the *ex vivo* anatomical structure) and $\boldsymbol{F_{cmd}} \in \mathbb{R}^m$ denotes the commanded force to the robot.

B. Learning-based Hybrid Needle Tip Tracking Strategy

As depicted in Fig. 3, a learning-based hybrid needle tip tracking framework has been proposed and mainly comprises three key modules: (1) "*Template-based Detection*" module, (2) "*YOLO-based Detection*" module, (3-1)"*Bounding Box (BBOX) Selection*, and (3-2) "*Dynamic Template Update*" module. The real-time US image acquisition is achieved using the "*Image Acquisition*" module, as depicted in Fig. 3.

1) Template-based Detection Module: This module incorporates: (1) the *layer-wise* feature aggregation from the "*Feature Extractor Block*", which leverages hierarchical information across different convolution layers, depicted in Fig. 3(a); (2) the lightweight *depth-wise correlation* in the RPN block, as depicted in Fig. 4(b), employed to maintain computational efficiency and reduce processing overhead. Furthermore, an improved template-matching-based detection method based on the Siamese Region Proposal Network (RPN) [24] is employed to facilitate efficient localization, as illustrated in Fig. 3.

Image Input to the Module: As shown in Fig. 3, the "*Template-based Detection*" module takes the US images as input from three branches, including:

- A real-time US image from the "*Image Acquisition*" module, as depicted in the left side of Fig. 3. The images have been cropped to 271x271 pixels from the original US images, which were 640x640 pixels in size, serving as the search region for needle tip detection.
- A dynamic template representing the latest appearance of the needle tip in the US image context with a size of 127x127 pixels;

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

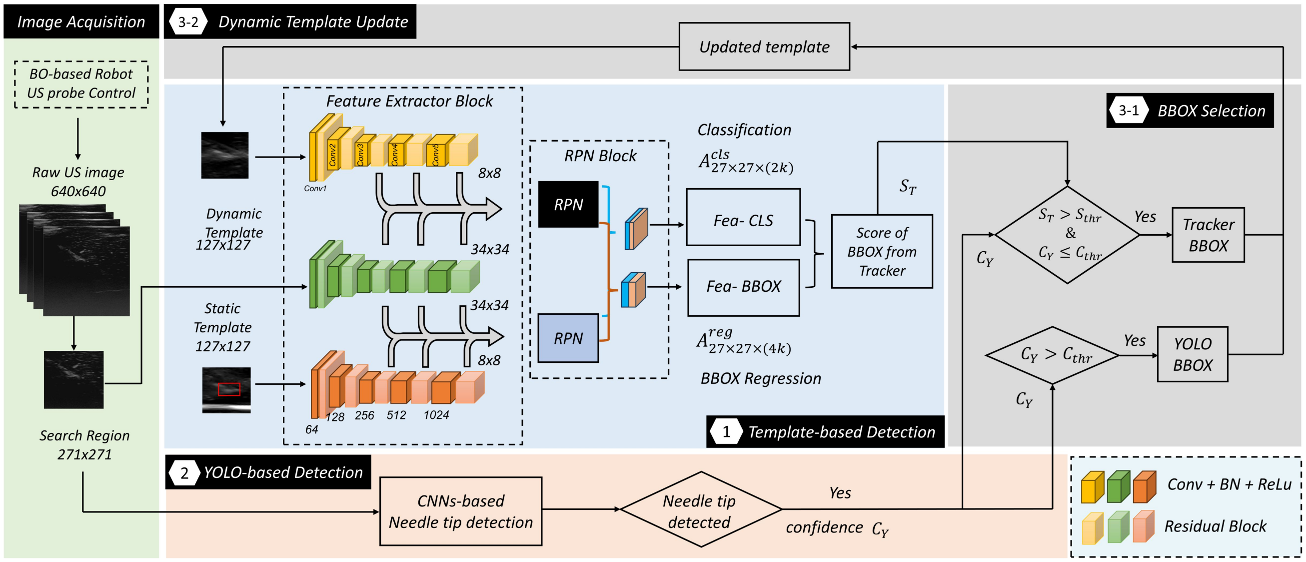

Fig. 3. Architecture of the proposed hybrid learning-based needle tip tracking framework. Specifically, the proposed framework consists of: (1) "*template-based detection*" module, (2) "*You Only Look Once (YOLO)-based detection*" module, and (3) "*bounding box selection and dynamic template update*" modules, which are indicated with cyan-blue, orange, and gray backgrounds, respectively. "*RPN*", "*Fea-CLS*", and "*Fea-BBOX*" are the Region Proposal Network, the feature classification, and the feature bounding box regression, respectively.

- A static template with a visible needle tip in the US image; (left side of the module, canny blue background).

Feature Extractor Block: In Fig. 3, a deep feature extractor block is first employed to extract rich and hierarchical features across different convolution layers from US images effectively. Relatively large kernels are utilized in the initial three convolution layers to provide a broader receptive field (9x9, 7x7, and 7x7, respectively). At the same time, down-sampling is achieved by setting the stride to 2. Furthermore, as depicted in Fig. 3, to merge the features from both the earlier and later layers, the outputs from the last three residual blocks are extracted and aggregated to infer the needle tip position collaboratively through a *layer-wise* feature fusion strategy in the feature extractor block. The number of these three channels is 256, 512, and 1024, respectively. Specifically, as depicted in Fig. 4 (a), each residual block, following a ResNet-like structure, incorporates a dilation convolution layer with a 3×3 kernel and a stride of 1. Additionally, a 1×1 convolution is applied in the shortcut path to match dimensions. Each residual block also integrates a Layer Normalization (LN) and a Gaussian Error Linear Unit (GeLU) activation function to enhance training stability and smooth convergence, as demonstrated in the vision transformer structure [25].

Regional Proposal Network (RPN) Block: As illustrated in Fig. 3, the RPN block is leveraged to perform foreground-background classification and anchor-based proposal refinement, i.e., needle tip detection and bounding box generation. Specifically, the parameter k in the "*Template-based Detection*" module indicates the number of anchor candidates associated with each pixel in the final feature maps of the US image. Moreover, each RPN block receives input from the final three residual blocks of the "*Feature Extractor Block*" in *"Template-based Detection"* module of Fig. 3.

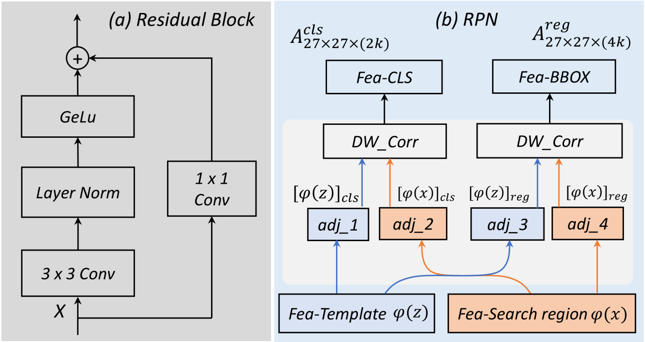

Fig. 4. Details of the *Residual* and *RPN blocks*. "*GeLu*", "*adj_i*", "*Fea*", and "*DW_Corr*" represent the Gaussian error Linear Unit, adjustment, feature, and Depth-Wise Correlation, respectively.

In Fig. 4 (b) of the RPN block, the features for correlation are firstly downed to a lower number of channels using the *"adj_i"* (i = 1, 2, 3, 4) convolutional block and made it suitable for tracking tasks. Afterward, the "*DW_Corr*" block layer predicts the correlation features between the templates and search regions in a multi-channel manner [26]. Specifically, the correlation process is expressed as:

$$\begin{cases} A^{cls}_{27\times27\times(2k)} = [\varphi(x)]_{cls} \star [\varphi(z)]_{cls} \\ A^{reg}_{27\times27\times(4k)} = [\varphi(x)]_{reg} \star [\varphi(z)]_{reg} \end{cases} \tag{6}$$

where $[\varphi(z)]_{cls}$, $[\varphi(z)]_{reg}$, $[\varphi(x)]_{cls}$, $[\varphi(x)]_{reg}$ denote the template feature maps and search region feature maps for classification and regression, respectively. $\star$ represents the convolution operation. $A^{cls}_{27\times27\times(2k)}$ saves the negative and positive activation of each anchor for classification, while $A^{reg}_{27\times27\times(4k)}$ denotes the distance between the anchor and the corresponding ground truth in the image.

Ultimately, as depicted in Fig. 3 and Fig. 4 (b), the *Fea-CLS* in the RPN block enables the network to predict whether each

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

anchor contains the needle tip (foreground) or belongs to the background, with size $A^{cls}_{27\times27\times(2k)}$, which saves the negative and positive value of the classification. In addition, the *Fea-BBOX* is used to predict translations relative to the offsets with size of $A^{reg}_{27\times27\times(4k)}$, which will be used to calculate the bounding box position afterwards. The final output of this module is the bounding box with the highest score, S_T, within the search region of the US image, representing the detected position of the needle tip.

2) YOLO-based Needle Tip Detection Module: Although template-matching-based approaches have demonstrated accuracy and efficiency, they are susceptible to mis-detection when image noise closely resembles the target. This issue is particularly critical for needle tip tracking tasks in the context of US images, where speckle patterns from anatomical structures and tissues often resemble the needle tip [3]. To overcome this limitation, a CNN-based detection approach, specifically the YOLO-based detection architecture, is integrated to facilitate single-stage object detection directly from the whole US images, which effectively mitigates the misidentification of speckle artifacts as the needle tip, a common issue associated with traditional template-matching-based methods. In addition, YOLO is first adopted to perform the static template initialization in the "*template-based detection*" module as well.

In Fig. 3, the YOLO-based model outputs the coordinates of the bounding box (upper-right and lower-left corners), the confidence score of the detection C_Y, and the predicted class label (i.e., needle tip or background). Considering that YOLO performs detection across the entire US image, the output can be directly considered as the detection results when the value of C_Y exceeds a confidence threshold (C_{thr}).

3) BBox Selection and Dynamic Template Update Module: To address the limitations of single-method needle tip detection in ultrasound (US) images, a comprehensive hybrid strategy has been developed to enhance efficiency and robustness. The proposed framework comprises two main components:

- 3.1) A "*BBOX Selection*" module, which identifies the needle tip location within the US image using a bounding box. As illustrated in Fig. 3, this module integrates confidence scores from two sources: the "*Template-based Detection*" (S_T) and "*YOLO-based Detection*" (C_Y). The results from the "*Template-based Detection*" serve as the primary detection method, while the "*YOLO-based Detection*" module supplements it when S_T falls below a predefined threshold (S_{thr}).
- 3.2) A "*Dynamic Template Update*" module, which adaptively refines the needle tip's appearance representation to accommodate variations in tissue structure and image appearance. This is achieved by cropping the region corresponding to the BBOX in the current frame and using it as the updated template.

C. Bayesian Optimization for Needle Tip Visualization Restore

Considering the challenge of needle tip invisibility resulting from unexpected movements, this section introduces a Bayesian Optimization (BO)-based robotic strategy for active US probe pose adjustment. The proposed approach aims to efficiently restore the needle tip's visibility when it disappears from the US image, eliminating the need for exhaustive frame-by-frame scanning of the surrounding area.

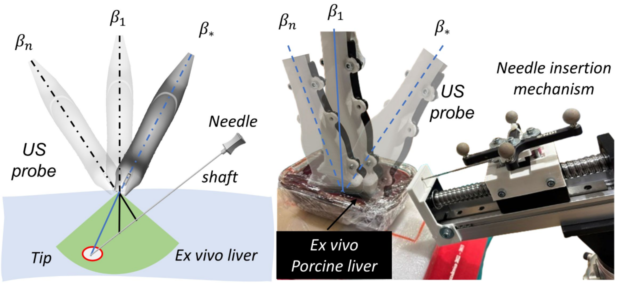

Fig. 5. The US probe tilting pose optimization for needle tip re-localization within the searching space. Left: illustration of the needle tip localization by adjusting the probe pose, β; Right: the implementation on the robot.

Notably, in this work, the needle tip is assumed to remain within the surrounding anatomical regions when it becomes invisible. As depicted in Fig. 5, the typical approach is to adjust the US probe pose, denoted by β, relative to its initial pose within a designated search region $\mathcal{B}$. The relationship between the needle tip detection results (confidence score) and the probe pose β is typically complex and not explicitly known. Specifically, the unknown function $f(\beta)$ that gives the confidence value of the needle tip detection at the US probe pose β is formulated in this work, which is expressed as below:

$$\max_{\beta\in\mathcal{B}} f(\beta) \tag{7}$$

To solve the optimization problem expressed by Eq. 7 for needle tip re-localization, the Bayesian Optimization, a numerical method well-suited for optimizing expensive black-box functions, is employed [27]. Specifically, it comprises three main components: a surrogate model, an acquisition function, and a stopping criterion. The detailed steps for adjusting the US probe pose to relocate the needle tip position utilizing Bayesian Optimization are summarized in the Algorithm 1, which consists of the following steps.

Algorithm 1 Bayesian Optimization for US Probe Pose β in Needle Tip Visualization Restore.

```
Input: Initial observation dataset, $\mathcal{D} = \{\beta, y\}$; Search area, $\mathcal{B}$; Maximum iterations, $N_{\max}$.
while Iteration $i \le N_{\max}$ do
    $\mathcal{GP}(\mu, \sigma^2) \leftarrow \mu, \sigma^2$
    $\beta_* \leftarrow \alpha_{\mathrm{EI}}(\beta; \mathcal{D})$
    $C_o \leftarrow f(\beta_*)$
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\beta_*, f(\beta_*)\}$
end while
return $\mathcal{D}$
Optimal US probe pose: $\beta^+ \leftarrow \arg\max f(\beta), \beta \in \mathcal{D}$
Output: Detected needle tip position, $P^b_{tip}$; US probe pose, $\beta^+$.
```

Step 1: In this work, a Gaussian Process (GP) is employed as the surrogate model to approximate an unknown function,

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

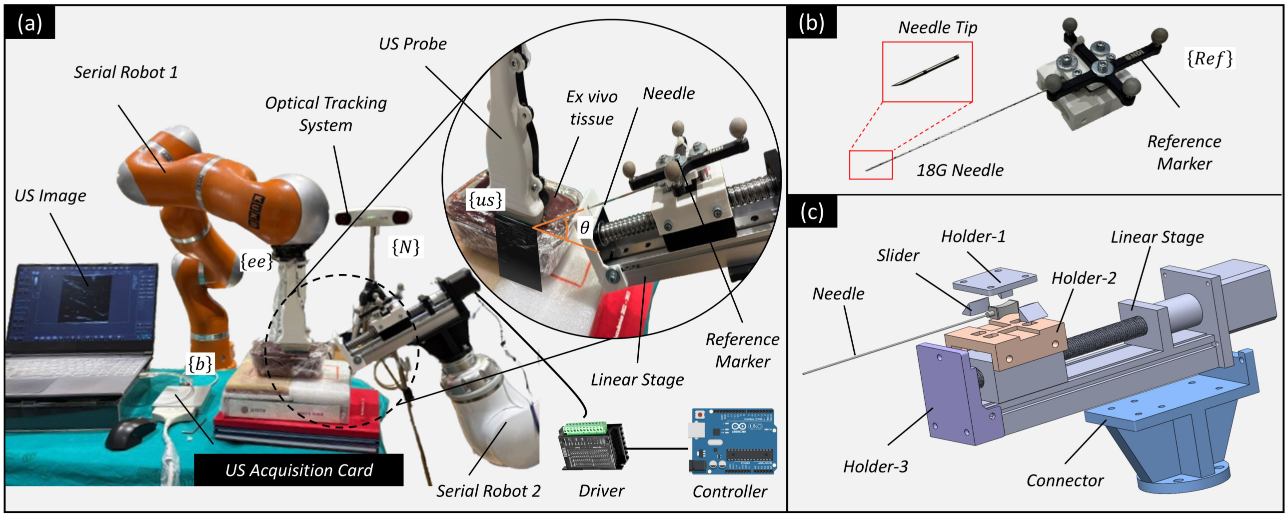

Fig. 6. Details of the proposed system for robotic Ultrasound-guided interventions. (a) depicts the overall system setup, includes a "*Serial Robot 1*" that holds the US probe for real-time image acquisition, while "*Serial Robot 2*" is leveraged to hold the linear stage module that performs automatic needle insertion, and the optical tracking system is considered as the ground-truth for error measurement. θ represents the needle insertion angle with respect to the "*Serial Robot 1*" base frame. (b) presents the detailed structure of the biopsy needle with a reference marker attached. (c) illustrates the detailed structure of the "*needle insertion module*" attached to the "*Serial Robot 2*".

f, using a prior mean function, m_f, and a covariance (kernel) function, k_f, formally expressed as:

$$f(\beta) \sim \mathcal{GP}(m_f, k_f) \tag{8}$$

Specifically, as depicted in Fig. 5, given an initial dataset $\mathcal{D} = \{\beta, y\}$, where $\beta = [\beta_1, \beta_2, \ldots, \beta_n]$ represents a set of initial US probe poses, and $y = [f(\beta_1), f(\beta_2), \ldots, f(\beta_n)]$ corresponds to the needle tip detection results with associated confidence values at these poses. Then, the GP is utilized to model the function, f, based on $\mathcal{D}$. Subsequently, for a new query the US probe poses, β_*, the posterior distribution conditioned on the dataset $\mathcal{D}$ can be derived as follows:

$$\mathcal{P}\left(f\left(\beta_*\right) \mid \beta_*, \mathcal{D}\right) = \mathcal{N}\left(\mu\left(\beta_*\right), \sigma^2\left(\beta_*\right)\right) \tag{9}$$

where $\mu\left(\beta_*\right)$ and $\sigma^2\left(\beta_*\right)$ denote the posterior mean and covariance functions, respectively. Specifically, $\mu\left(\beta_*\right) = m_f\left(\beta_*\right) + k_* K^{-1}\left(y - m_f(\beta)\right)$, $\sigma^2\left(\beta_*\right) = k\left(\beta_*, \beta_*\right) - k_* K^{-1} {k_*}^T$, and $k_* = \left[k\left(\beta_*, \beta_1\right), \ldots, k\left(\beta_*, \beta_n\right)\right]^T$. Considering the robustness, flexibility, and performance on rough functions, the Matérn kernel function with a smoothness parameter 5/2 is utilized as the kernel function, k, instead of using the squared exponential function [28]. Furthermore, K is calculated as:

$$K = \begin{bmatrix} k\left(\beta_1, \beta_1\right) & \cdots & k\left(\beta_1, \beta_n\right) \\ \vdots & \ddots & \vdots \\ k\left(\beta_n, \beta_1\right) & \cdots & k\left(\beta_n, \beta_n\right) \end{bmatrix} \tag{10}$$

Step 2: After modeling the objective function f using Gaussian Process (GP) regression as a surrogate model, the subsequent step is to infer the next exploration probe pose, β_*, for US image acquisition and needle tip detection. In each BO iteration, β_* is determined to maximize the probability of locating the needle tip can be formulated as follows:

$$\beta_* = \arg\max_{\beta \in \mathcal{B}} \{\alpha(\beta; \mathcal{D})\} \tag{11}$$

where α denotes the acquisition function for determining the next probe pose β_* for exploration. Specifically, the Expected Improvement (EI) function, α_{EI}, is utilized as the acquisition function α in Eq.(11), which is expressed by:

$$\alpha_{EI} = \begin{cases} \left(\mu(\beta) - f^+ - \zeta\right)\phi(Z) + \sigma(\beta)\varphi(Z), \sigma(\beta) > 0 \\ 0, \text{ otherwise} \end{cases} \tag{12}$$

where $Z = \frac{\mu_f(\beta) - f^+ - \zeta}{\sigma(\beta)}$. Φ and φ denote the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution, respectively. The posterior mean and covariance based on the dataset $\mathcal{D}$ are represented by μ and σ^2. The term f^+ refers to the best observed value so far. Moreover, ζ controls the balance between exploration and exploitation during the optimization using the Expected Improvement (EI) criterion [29].

Step 3: Upon obtaining the observation of the new US probe pose β_* with the results $\{\beta_*, f(\beta_*)\}$, the augmented dataset $\mathcal{D}$ for updating the GP model can be expressed as $\mathcal{D} \leftarrow \mathcal{D} \cup \{\beta_*, f(\beta_*)\}$. Specifically, the optimization process proceeds until one of two termination conditions is met: either the maximum number of interactions, $N_{\max}$, has been reached, or the confidence value for needle tip detection exceeds a predefined threshold. To improve the robustness, the confidence value, C_o, for the US probe pose optimization is computed using the formula below:

$$C_o = f(\beta) = \gamma S_T + (1 - \gamma) C_Y \tag{13}$$

where γ is a weighting factor to balance the contributions from the two detection modules in Fig. 3.

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

III. EXPERIMENTAL SETUP

A. System Implementation

The detailed components of the developed system for robotic US-guided needle interventions are illustrated in Fig. 6. They are primarily composed of the following elements:

- A 7-DOFs serial robot 1, LWR IV+ (KUKA Roboter GmbH, Germany), shown in Fig. 6 (a). This robot incorporates the FastResearchInterface (FRI) and the Robot Operating System (ROS), and is responsible for manipulating the US probe [30].
- A linear ultrasound probe (L12-5L40S-3, Telemed, Vilnius, Lithuania) mounted on the end effector of the robotic arm to facilitate US imaging.
- A frame grabber (Magewell Gen2, China) employed to capture raw US images from the acquisition card, enabling synchronized image transfer to a desktop computer for subsequent image processing and needle tip detection at a frame rate of 30 *fps*.
- In Fig. 6 (b), an 18-gauge biopsy needle, 150 mm in length and 1.2 mm in diameter, is employed for the ultrasound-guided needle insertion procedure. The flexibility of this needle is considered negligible due to its inherent rigidity [3], [19]. Additionally, the component labeled "*Holder 3*" in Fig. 6 (c) serves to minimize needle deflection during the insertion procedure.
- Furthermore, an external optical tracking system, the Polaris Vicra (NDI, Northern Digital Inc., Canada), was employed to localize and record the needle tip's position for subsequent validation and comparative experiments.

Furthermore, to evaluate needle tip tracking performance across varying needle insertion velocities and angles, a customized *needle insertion module* was developed to ensure standardization, as depicted in Fig.6. For precise positioning of the module, another 7-DOFs robotic arm, the LBR iiwa 7 (KUKA Roboter GmbH, Germany), denoted as Serial Robot 2 in Fig. 6 (a) and controlled with FRI, incorporates ROS2 [1] , was employed to manipulate the needle insertion module via a designed connector, as shown in Fig. 6 (c). Specifically, the module integrates a linear stage with programmable speed control, driven by a TB6600 motor driver and an Arduino UNO microcontroller. This setup facilitates consistent and precise needle insertions, thereby enhancing the reliability of the results under controlled experimental conditions.

B. Dataset Collection, Training, and Testing

To build the dataset for training the network, eight fresh *ex vivo* porcine liver samples were used. Specifically, as depicted in Fig. 6 (a), each sample was placed in a container with an approximate tissue depth of 60 mm. Notably, the needle tip didn't reach the bottom during the whole procedure. US imaging videos were captured using a frame grabber connected to the US imaging machine, with raw US image frames resized to 640×640 pixels and recorded at 30 frames per second. To replicate the variability encountered in clinical procedures such as biopsies and regional anesthesia, the dataset was designed to encompass a wide range of insertion conditions. Specifically, as depicted in Fig. 7, variations in needle insertion angles (approximately 10°–45°), depths (10-40 mm), and velocities (0.5-5 mm/s) were systematically introduced. A total of 50 videos were recorded to build the dataset, each capturing manual needle insertion and withdrawal maneuvers. It should be noted that the deflection of the needle tip during the needle-tissue interaction is not considered. From these recordings, approximately 11,000 US frames were extracted following further refinement. Afterwards, the ground truth annotations of the needle tip's position within the US image were manually implemented using the online annotation platform Roboflow, with high-precision, pixel-level accuracy of annotation.

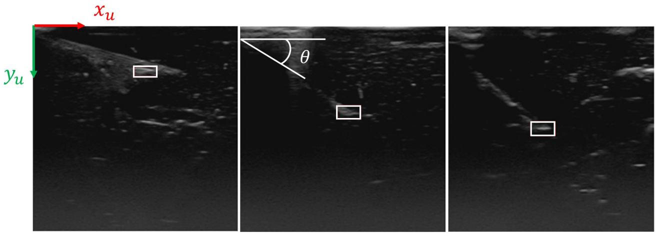

Fig. 7. Illustration of the US images during the needle tip insertion into the *ex vivo* porcine liver with different angle θ.

The training of the proposed framework was implemented with 50 epochs, a batch size of 16. The Stochastic Gradient Descent (SGD) optimizer was utilized during the training process, with the learning rate of an initial value of 1×10^{-5}, a weight decay factor 5×10^{-4}, and 0.9 was set for a momentum in SGD. The composite loss function integrated a Binary Cross-Entropy (BCE) loss for classification and a weighted L_1 loss for bounding box regression, denoted as L_{cls} and L_{reg}, respectively [24]. Specifically, to balance the localization and class prediction accuracy, the loss function, L_{loss}, is formulated as below:

$$L_{loss} = \lambda L_{cls} + (1-\lambda) L_{reg} \tag{14}$$

where λ is a hyperparameter to balance these two parts. In this work, YOLOv11 [31], a pre-trained CNN-based architecture, was employed to perform the needle tip detection. The training of the YOLO-based detection module was implemented on the same dataset. Specifically, the dataset was systematically divided into training and validation subsets with a ratio of 8:2. All the training was conducted on a workstation equipped with an Intel Core i7-14700KF CPU (3.40 GHz, 32 GB RAM) and an Nvidia GeForce RTX 4060 Ti GPU (16 GB VRAM) to leverage parallelized computation during training.

C. Experimental Protocol

To investigate the performance of the proposed needle tip tracking framework, a series of experiments has been designed and implemented. Specifically, Experiment 1 (E1): *"Needle Tip Tracking Performance Validation Experiment"* and Experiment 2 (E2): *"Needle Tip Visualization Restore Validation Experiment"* have been designed.

Experiment 1: Needle Tip Tracking Performance Validation. This experiment aims to implement a comprehensive

---

[1] https://github.com/lbr-stack

verification of the needle tip tracking performance with the proposed hybrid tracking framework in Section II-B. As illustrated in Fig. 7, the appearance of the needle varies under different insertion angles, along with artifacts, speckles, and reflections. Specifically, as depicted in Fig.7, the needle is inserted from the upper left to the right side, with an insertion distance of approximately 50 mm, followed by a withdrawal procedure. To further investigate the tracking performance, three needle insertion velocities $v = [0.5, 1.5, 3.0]$ mm/s and three insertion angles $\theta = [15°, 30°, 40°]$ are investigated, resulting in nine different experimental setups [32]. The needle insertion angle, θ, and velocity, v, are precisely controlled utilizing the Serial Robot 2 and the custom-designed needle insertion module, as depicted in Fig. 6 (c).

The real-time needle tip tracking experiment is first implemented with the proposed framework. Specifically, the initialization of the "*Template*" within the "*Template-based Detection*" module in Fig.3 is extracted from the YOLO-based detection results with the highest confidence value under static conditions. To address the morphological variability of needle tip appearances in US images, anchor-based detection parameters were refined with a size of 8, and five aspect ratios $[0.3, 0.5, 1, 2, 3]$ were selected to accommodate scale and diversity. An external optical tracking system was fixed next to the robotic manipulator to record the ground truth position during the experiment, as depicted in Fig.6 (b). Moreover, to ensure consistency, each experiment scenario is conducted with nine repetitions to evaluate the needle tip tracking results.

Furthermore, comparison experiments have been conducted to verify the needle tip tracking performance of the proposed hybrid tracking framework compared to the YOLOv11 [31] and SiamRPN++ [26], [33] framework-based detection and tracking approaches. The comparison of needle tip detection and tracking results across the three tracking approaches was computed offline, utilizing recorded ultrasound video sequences and corresponding ground-truth needle tip positions during the previous experiments for fair comparison.

Experiment 2: Needle Tip Visualization Restore Validation This experiment evaluates the effectiveness of the proposed BO-based strategy in restoring needle tip visibility. To quantify the proposed needle tip visualization performance, several standardized steps are implemented to mimic the scenarios where the needle tip becomes invisible within the US image frames, as depicted in Fig. 8.

- In *Step 1*, the needle is initially inserted into the *ex vivo* porcine liver tissue using the motorized insertion module until the tip reaches the center of the US image.
- Subsequently, in *Step 2* of Fig. 8, the US probe is controlled to rotate around the z-axis, representing the misalignment angle, α, between the US imaging plane and the needle shaft.
- Following this, the needle is further advanced by the needle insertion mechanism in *Step 3* with a displacement d, making the needle tip invisible within the US image, as depicted in column 3 of Fig. 8.

The BO-based robotic US probe pose adjustment procedure, described in Section II-C, is initiated to relocate the needle tip into the imaging plane. The US probe roll angle, β, is empirically constrained within the range of –0.3 to 0.3 rad relative to its initial roll orientation. The optimization process begins with eight randomly selected initial sampling poses to construct the dataset $\mathcal{D}$, and the exploration steps $N_{\max}$ is empirically set as 7. Furthermore, to assess the robustness of the needle tip relocation method, experiments are conducted using three distinct misalignment angles ($\alpha = 10°, 15°$, and $20°$), with each condition repeated three times.

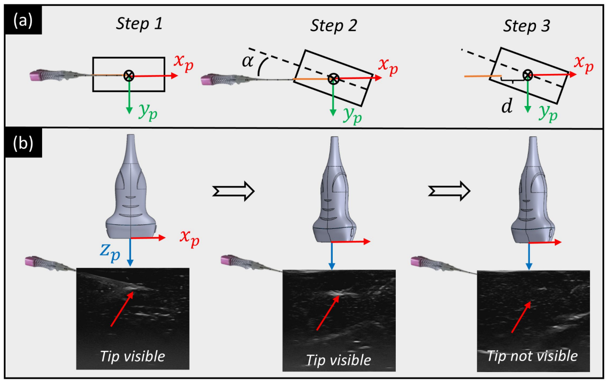

Fig. 8. Steps to mimic the needle tip disappear within the US image. (a) represents the sketch of the relative pose between the probe and needle from ***top view***. ⊗ represents the middle of the US image plane. $\boldsymbol{d}$ is the displacement of needle advancements; (b) the needle tip is not visible within the US image.

D. Performance Metrics

To evaluate the effectiveness of the proposed framework for needle tip tracking and visualization restoration, the Cartesian error, e_{tip} $[mm]$ in needle tip detection and localization, is computed. This metric quantifies the Euclidean distance between the ground truth needle tip position P^b_{tip}, which is measured by an external tracking system, and the position from the US image. Specifically, e_{tip} is calculated as:

$$e_{tip} = ||P^b_{tip} - P'^b_{tip}|| = ||{}^b_N\mathbf{T}P^N_{tip} - {}^b_{us}\mathbf{T}P^{us}_{tip}|| \qquad (15)$$

where $P^b_{tip} = {}^b_N\mathbf{T}P^N_{tip}$ and $P'^b_{tip} = {}^b_{us}\mathbf{T}P^{us}_{tip}$. ${}^b_N\mathbf{T} \in \mathbb{R}^{4\times4}$ represents the coordinate transformation from the optical tracking system to the robot base, which is obtained through hand-eye calibration, as implemented in [21]. $P^N_{tip} \in \mathbb{R}^3$ denotes the needle tip position in the frame of the optical tracking system, as illustrated in Fig. 6.

To analyze the results, the Kolmogorov-Smirnov-Lilliefors test is employed first to assess the data distribution's normality. The median (MED) and inter-quartile range (IQR) values of the needle tip tracking errors E_{med} and E_{iqr} are evaluated when the data are not normally distributed. Furthermore, the statistical analysis is performed utilizing the Wilcoxon Rank-Sum test. Specifically, a significant difference is observed with a $p-value < 0.05$, ($-$, $p \geq 0.05$ denotes no significant difference exists. $*$, $p < 0.05$; $**$, $p < 0.01$; $***$, $p < 0.001$).

IV. EXPERIMENT RESULTS

A. E1 Results: Needle Tip Tracking Performance Validation

1) Real-time Needle Tip Tracking Results with Proposed Framework: First of all, the real-time needle tip tracking

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

TABLE I
ONLINE NEEDLE TIP TRACKING ERROR, e_{tip}, WITH PROPOSED FRAMEWORK, INCLUDING THE MEDIAN ERROR, E_{med}, AND IQR, E_{iqr}.

Insertion velocity	Needle insertion angle θ		
	$\theta = 15°$	$\theta = 30°$	$\theta = 40°$
0.5 mm/s	1.03 (0.71)	0.54 (0.25)	0.86 (0.79)
1.5 mm/s	1.13 (0.92)	0.64 (0.34)	1.21 (0.95)
3.0 mm/s	1.28 (1.25)	0.79 (0.44)	1.25 (0.94)

results using our proposed hybrid tracking framework is presented in Fig. 9, with needle tip tracking error, e_{tip}, recorded during needle insertion and withdrawal phases across various insertion velocities and angles described in Section III-C. Moreover, Table. I summarizes the median error, E_{med}, and IQR value, E_{iqr} of the tracking results.

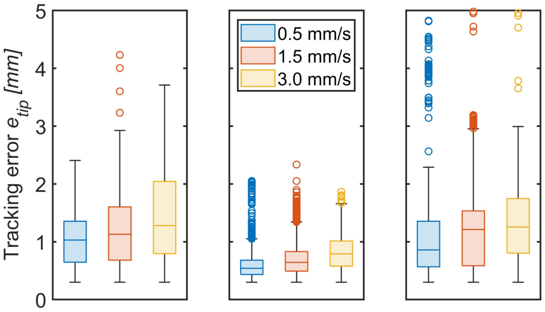

Fig. 9. Online needle tip tracking results with the proposed framework under different insertion speeds and angles. From left to right panels are the results with needle insertion angles of 15°, 30°, and 40°, respectively.

Specifically, as depicted in the Table. I and Fig. 9, lower needle insertion angles generally produce stronger echo signals, thereby improving needle tip visibility in US images. Notably, the online tracking experiment results with the insertion angle of $\theta = 30°$ in Fig. 9 achieve the lowest tracking error among the three needle insertion angles. However, at insertion angle $\theta = 15°$, as shown in Fig. 10 (row 1), reverberation artifacts, caused by strong specular reflections between the needle shaft and the probe, create a bright reflective area beneath the needle, leading to a slightly higher tracking error e_{tip}. Conversely, at $\theta = 40°$ (rows 2 and 3), weaker echo signals and increased artifacts and speckles diminish needle tip visibility, degrading the accuracy of bounding box predictions for tip detection and tracking. Furthermore, both Fig. 9 and Table. I indicates that e_{tip} increases consistently across all tested velocities when the insertion velocity increases. This degradation in tracking accuracy is primarily attributed to intensified needle–tissue interactions at higher speeds, which disrupt surrounding tissue structures and introduce additional noise and artifacts into the US images. Overall, the proposed method achieves a maximum median error, E_{med}, below 1.30 mm, demonstrating reliable overall tracking performance.

2) Comparison of Needle Tip Tracking Performance:

Moreover, comparative experiments were conducted across all experimental setups to evaluate the proposed hybrid needle

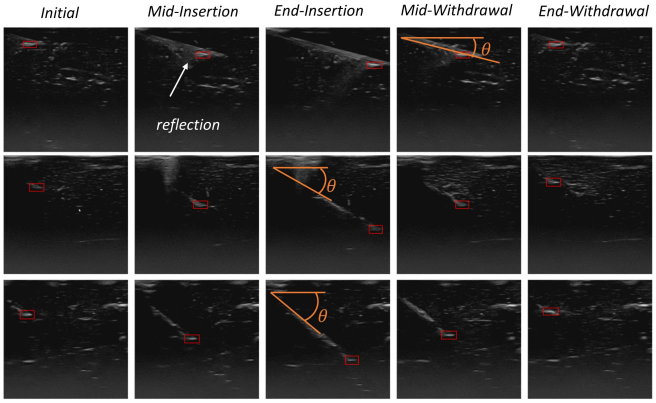

Fig. 10. Needle tip appearance within the US image and detected needle tip. From top to bottom rows are the insertion angles θ with 15°, 30°, and 40°, respectively. From left to right are the *"Initial"* (column 1), *"insertion"* (columns 2 and 3), and the *"withdrawal"* procedures (columns 4 and 5).

tip detection and tracking framework against two state-of-the-art detection and tracking architectures: YOLOv11 [31] and SiamRPN++ [26], [33], as described in Section.III-C. The median errors, E_{med}, and IQR values, E_{iqr}, of needle tip tracking error, e_{tip}, for each method across all the setups are summarized in the Table. II. Specifically, for the YOLO-based modality, the best performance was observed at an insertion angle of 15° and an insertion velocity of 0.5 mm/s, achieving a median error of 0.90 mm. This indicates that YOLO-based approaches can be effective for needle tip detection and tracking within the US images, however, only under favorable conditions. Subsequently, as both insertion angle and velocity increase, the median error (E_{med}) and IQR (E_{iqr}) also increase. As illustrated in Fig. 11, in the case of insertion angle at 40° with a velocity of 3.0 mm/s, the YOLO-based method fails to reliably detect the needle tip, resulting in significant fluctuations in tracking error. This degradation in performance is likely due to reduced needle tip visibility in US images at steeper angles and the inherent limitations of YOLO in detecting small or closely adjacent objects.

Furthermore, the second row in the Table. II presents the results for the SiamRPN++ approach, which generally outperforms the YOLO-based method. Notably, it achieves the best tracking accuracy at an insertion angle of 30° across all tested velocities, outperforming the 15° and 40° configurations. These results indicate that SiamRPN++ is more robust across varying velocities and better suited for detecting small objects compared to the YOLO-based one. The lowest median error of 1.26 mm at an insertion angle of 15° and a velocity of 3.0mm/s. The proposed hybrid tracking framework demonstrates superior performance over both baseline methods. Specifically, as detailed in Table II, it achieves the lowest median error of 0.53 mm at an insertion angle of 30° and a velocity of 0.5mm/s. Across all the experiment setups, the highest observed median error was 1.20 mm, occurring at an insertion angle of 40° and a velocity of 3.0mm/s.

Notably, as depicted in Fig. 11, the proposed tracking framework demonstrates the capability to reliably track the needle tip even in frames where SiamRPN++ fails to main-

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

TABLE II
COMPARISON OF NEEDLE TIP TRACKING ERROR, e_{tip}, WITH DIFFERENT APPROACHES. (MEDIAN ERROR, E_{med}, AND IQR, E_{iqr}, $[mm]$)

Setups	Insertion velocity 0.5 mm/s			Insertion velocity 1.5 mm/s			Insertion velocity 3.0 mm/s		
	15°	30°	40°	15°	30°	40°	15°	30°	40°
YOLOv11	0.90 (0.66)	12.20 (20.04)	14.46 (13.91)	1.12 (1.00)	12.30 (12.45)	13.14 (24.54)	1.33 (1.23)	11.45 (32.99)	18.37 (19.27)
SiamRPN++	0.93 (0.73)	0.53 (0.28)	0.91 (0.86)	1.09 (0.98)	0.83 (0.73)	1.33 (0.99)	1.26 (1.11)	0.66 (0.43)	1.22 (1.12)
Hybrid Tracking	0.97 (0.72)	0.53 (0.22)	0.89 (0.78)	1.09 (0.88)	0.66 (0.30)	1.20 (0.93)	1.17 (1.28)	0.56 (0.29)	1.09 (1.10)

TABLE III
STATISTICAL ANALYSIS OF THE COMPARISON FOR NEEDLE TIP TRACKING ERROR, e_{tip}.

Comparison Setups	Velocity at 0.5 mm/s			Velocity at 1.5 mm/s			Velocity at 3.0 mm/s		
	15°	30°	40°	15°	30°	40°	15°	30°	40°
SiamRPN++ *vs.* YOLOv11	*	* * *	* * *	**	* * *	* * *	* * *	* * *	* * *
Hybrid Tra.(Proposed) *vs.* YOLOv11	* * *	* * *	* * *	* * *	* * *	* * *	* * *	* * *	* * *
Hybrid Tra.(Proposed) *vs.* SiamRPN++	–	*	* * *	–	* * *	* * **	–	* * *	* * *

tain accurate tracking. However, SiamRPN++ occasionally misidentifies speckle artifacts within the US image as the needle tip due to their visual similarity. When this occurs, the tracking error increases significantly, and the method fails to recover the correct needle tip position, which is an inherent limitation of the template-matching mechanism in the SiamRPN++ architecture. Further analysis of the tracking error curves from all these 81 trials revealed that the YOLO-based method demonstrated frequent tracking failures, with errors occurring even at a 15° insertion angle and increasing in frequency with higher insertion velocities. The SiameRPN++ method exhibited 35 instances where the tracking error exceeded 2 mm, predominantly at a 40° insertion angle and a velocity of 1.5 mm/s. In comparison, the proposed hybrid tracking framework showed improved robustness, with only 17 occurrences of errors greater than 2 mm, though similar trends of increased error were observed when the insertion angle and velocity increased. Overall, the results from Table II and Fig. 11 show that the hybrid framework maintains accurate and stable needle tip tracking, even in the presence of speckle noise. The maximum median value of the tracking error remains below 2 mm, underscoring the framework's effectiveness.

In addition, to assess whether significant differences exist among the results obtained from different tracking modalities, statistical analysis was conducted using the Wilcoxon rank-sum test, with the corresponding p-values reported in Table III. As shown in rows 2 and 3, both the SiamRPN++ and the proposed hybrid tracking approach outperform the YOLO-based method across all configurations, exhibiting statistically significant differences. Furthermore, row 3 presents a comparison between SiamRPN++ and the proposed method. Notably, at a needle insertion angle of 15°, both approaches yield low tracking errors, and the difference between them is not statistically significant. However, at insertion angles of 30° and 40°, significant differences are observed between the two methods. As illustrated in Fig. 9, higher insertion angles lead to weaker needle tip appearances and echo signals in the ultrasound images, accompanied by increased artifacts and speckle noise. These effects reduce needle tip visibility and subsequently degrade the accuracy of bounding box predictions for tip detection and tracking.

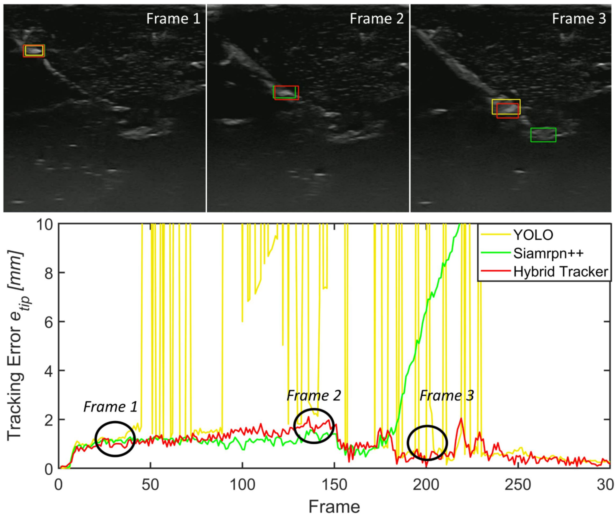

Fig. 11. Example of needle tip tracking with insertion speed at 3 mm/s and angle 40°. Row 1 depicts the needle tip detection results highlighted with bounding boxes in three different colors at three representative Ultrasound image frames; Row 2 gives the results of real-time needle tip tracking error measured by the external optical tracking system.

B. E2 Results: Needle Tip Visualization Restore Validation

To ensure the consistency of the needle tip visualization restoration experiments, the needle tip disappearance within the US image procedures is repeated independently after each trial, following the steps illustrated in Fig. 8 and described in Section III-C. Fig.12 presents the boxplot results of the needle tip repositioning error across three misalignment angles α with the proposed BO-based strategy. Specifically, based on statistical analysis of the restoration experiments, when the

This article has been accepted for publication in IEEE Transactions on Medical Robotics and Bionics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2026.3654277

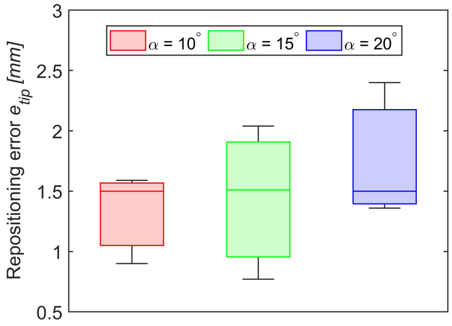

Fig. 12. Boxplot of the needle tip repositioning errors with BO-based strategy.

misalignment angle α is 10°, the median and IQR values of the needle tip relocation error (e_{tip}) are measured as 1.50 mm and 0.52 mm, respectively, representing the lowest localization error observed across all tests in the experiment. As the misalignment angle increases to 15° and 20°, the proposed optimization-based strategy continues to successfully visualize the needle tip within the US image, yielding localization errors of 1.52 (0.95) mm and 1.50 (0.78) mm, respectively. While these IQR values are marginally higher than those at 10°, no significant differences are observed according to the statistical analysis of the results.

Furthermore, Fig. 8 illustrates the procedure for adjusting the US probe pose using the proposed optimization-based strategy to restore needle tip visualization within the US imaging plane. Initially, a series of random probe poses (*S1* to *S8*), shown in Row 1, are sampled within a predefined search region $\mathcal{B}$ on an *ex vivo* porcine liver. These samples are used to construct the dataset $\mathcal{D}$ for training the Gaussian Process model. In Row 2, the needle tip detection results are presented with bounding boxes. However, due to the needle tip being outside the current US frame, the detection confidence is quite low (approximately 0.2), and the bounding boxes localize the speckles as the needle tip instead of the target. Following the initialization of $\mathcal{D}$, the proposed BO-based active exploration strategy is employed to infer the next optimal probe pose (denoted as *E1*) for robotic manipulator exploration. The corresponding roll angle, β, adjustments are illustrated in Row 3 of Fig. 13 during the exploration process. Notably, the needle tip is almost correctly identified in explorations *E1* and *E6*, however, with lower detection confidence compared to *E7*. In exploration *E7*, as illustrated in Row 4 of Fig. 13, the needle tip is successfully detected with a high confidence score of 0.43, and the localization error compared to the ground truth is below 1.50 mm. These results consistently demonstrate that the proposed optimization-based probe adjustment strategy is both effective and accurate in restoring needle tip visibility after performing targeted pose explorations.

V. CONCLUSION

This study presents a hybrid learning-based framework for needle tip tracking and visualization in robotic US-guided surgical interventions. Specifically, the proposed framework consists of a dual-branch template matching-based detection module, a CNN-based detection module, and a bounding box selection and template updating mechanism to enhance the accuracy and robustness of the needle tip tracking in US images. To further address the challenge of needle tip disappearance caused by potential motion artifacts, a BO-based recovery strategy is incorporated to guide the robot arm for active exploration. To validate the effectiveness of the proposed framework, experiments were first conducted on *ex vivo* porcine liver samples under varying needle insertion angles and velocities. Comparative experiments demonstrate that the proposed method achieves superior tracking performance, with a median localization error consistently below 1.30 mm. Additionally, in scenarios involving needle tip occlusion or disappearance, the BO-based restoration technique maintained localization errors within 1.52 mm across all tested conditions. Despite these promising results, the current framework has certain limitations. Firstly, it does not account for dynamic alignment between the US probe and the needle shaft, which requires additional information about the needle shaft detection and pose calculation. Moreover, the experimental validation was restricted to static *ex vivo* tissue samples, whereas clinical applications often involve physiological motions such as respiration [34], [35]. Future work will focus on integrating model predictive control schemes to compensate for the dynamic disturbances, thereby improving the clinical applicability. Furthermore, the deflection of the needle tip resulting from tip–tissue interaction requires further investigation and consideration [36], [37].

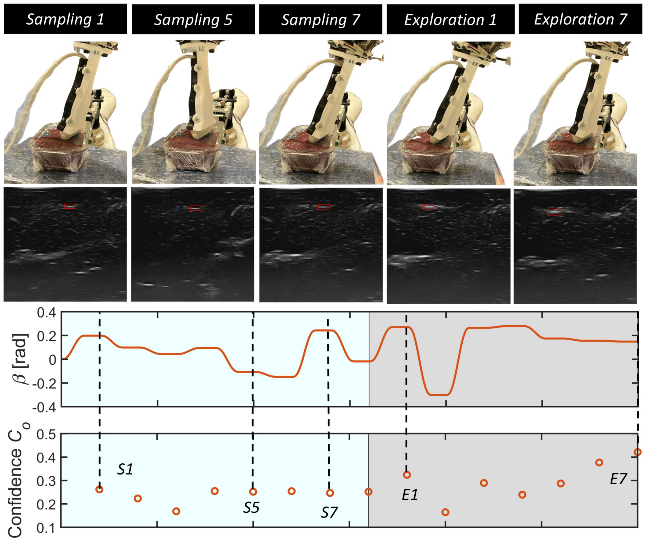

Fig. 13. Snapshots during probe adjustment for needle tip localization. Rows 1 to 4 are the US probe pose, US images, probe roll angle β, and the needle tip detection score C_o. Specifically, *S1*, *S5*, and *S7* represents samplings for initialization. *E1* and *E7* represent two explorations at the poses inferred by the BO-based optimization.

REFERENCES

[1] T. de Baere, C. Roux, G. Noel, A. Delpla, F. Deschamps, E. Varin, and L. Tselikas, "Robotic assistance for percutaneous needle insertion in the kidney: preclinical proof on a swine animal model," *European Radiology Experimental*, vol. 6, no. 1, p. 13, 2022.

[2] Z. Cheng, M. Koskinopoulou, S. Bano, D. Stoyanov, T. R. Savarimuthu, and L. S. Mattos, "Sensing technologies for guidance during needle-based interventions," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[3] D. Huang, C. Li, A. Karlas, X. Chu, K. S. Au, N. Navab, and Z. Jiang, "Vibnet: Vibration-boosted needle detection in ultrasound images," *IEEE Transactions on Medical Imaging*, 2025.

[4] Z. Liang, C. Lu, H. Yang, R. Hashem, M. E. Abdelaziz, L. Lindenroth, S. Bandula, D. Stoyanov, and A. Stilli, "An mr safe double-arch needle insertion robot with scissor-folding mechanism for abdominal percutaneous interventions," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11 010–11 017. IEEE, 2024.

[5] R. Li, A. Davoodi, M. Timmermans, K. Van Assche, O. Taylan, L. Scheys, M. Tummers, G. Borghesan, and E. Vander Poorten, "Ultrasound-based robot-assisted drilling for minimally invasive pedicle screw placement," *IEEE Transactions on Medical Robotics and Bionics*, vol. 6, no. 3, pp. 818–828, 2024.

[6] A. Rahaman, Y. Tang, S. Gao, X. Ma, I. Sorokin, and H. K. Zhang, "Needle aligned ultrasound image-guided access through dual-segment array," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 9, pp. 2645–2654, 2023.

[7] R. Moreta-Martínez, I. Rubio-Pérez, M. García-Sevilla, L. García-Elcano, and J. Pascau, "Evaluation of optical tracking and augmented reality for needle navigation in sacral nerve stimulation," *Computer methods and programs in biomedicine*, vol. 224, p. 106991, 2022.

[8] W. Zhang, S. Wang, L. Xie, and R. Yang, "An autonomous liver biopsy robotic system under respiratory motion," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[9] L. Grau-Mercier, A. Chetioui, L. Muller, C. Roger, R. Genre Grandpierre, J. E. de La Coussaye, P. Cuvillon, P.-G. Claret, and X. Bobbia, "Magnetic needle-tracking device for ultrasound guidance of radial artery puncture: A randomized study on a simulation model," *Journal of Clinical Ultrasound*, vol. 49, no. 3, pp. 212–217, 2021.

[10] Y. Lyu, R. Wang, J. Wang, L. Ji, and J. Cheng, "Multipole magnetic needle positioning method for ultrasound-guided puncture," *IEEE Transactions on Biomedical Engineering*, 2024.

[11] M. Kaya, E. Senel, A. Ahmad, and O. Bebek, "Visual needle tip tracking in 2d us guided robotic interventions," *Mechatronics*, vol. 57, pp. 129–139, 2019.

[12] M. Kaya and O. Bebek, "Needle localization using gabor filtering in 2d ultrasound images," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4881–4886. IEEE, 2014.

[13] A. Ayvaci, P. Yan, S. Xu, S. Soatto, and J. Kruecker, "Biopsy needle detection in transrectal ultrasound," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 653–659, 2011.

[14] K. Mathiassen, D. Dall'Alba, R. Muradore, P. Fiorini, and O. J. Elle, "Robust real-time needle tracking in 2-d ultrasound images using statistical filtering," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 966–978, 2016.

[15] J. A. Sánchez-Margallo, L. Tas, A. Moelker, J. J. van den Dobbelsteen, F. M. Sánchez-Margallo, T. Langø, T. van Walsum, and N. J. van de Berg, "Block-matching-based registration to evaluate ultrasound visibility of percutaneous needles in liver-mimicking phantoms," *Medical Physics*, vol. 48, no. 12, pp. 7602–7612, 2021.

[16] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Convolution neural networks for real-time needle detection and localization in 2d ultrasound," *International journal of computer assisted radiology and surgery*, vol. 13, pp. 647–657, 2018.

[17] S. Chen, Y. Lin, Z. Li, F. Wang, and Q. Cao, "Automatic and accurate needle detection in 2d ultrasound during robot-assisted needle insertion process," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2022.

[18] Y. Cai, R. Li, A. Davoodi, M. Ourak, J. Deprest, and E. Vander Poorten, "Autonomous robotic ultrasound approach for fetoscope tracking by fusing optical and 2d ultrasound data," *IEEE Robotics and Automation Letters*, 2024.

[19] Z. Jiang, X. Li, X. Chu, A. Karlas, Y. Bi, Y. Cheng, K. S. Au, and N. Navab, "Needle segmentation using gan: restoring thin instrument visibility in robotic ultrasound," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[20] C. Mwikirize, A. B. Kimbowa, S. Imanirakiza, A. Katumba, J. L. Nosher, and I. Hacihaliloglu, "Time-aware deep neural networks for needle tip localization in 2d ultrasound," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. 819–827, 2021.

[21] W. Yan, Q. Ding, J. Chen, K. Yan, R. S.-Y. Tang, and S. S. Cheng, "Learning-based needle tip tracking in 2d ultrasound by fusing visual tracking and motion prediction," *Medical Image Analysis*, vol. 88, p. 102847, 2023.

[22] D. Huang, C. Yang, M. Zhou, A. Karlas, N. Navab, and Z. Jiang, "Robot-assisted deep venous thrombosis ultrasound examination using virtual fixture," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 381–392, 2024.

[23] J. Fu, G. Maimone, E. Iovene, J. Zhao, A. Redaelli, G. Ferrigno, and E. De Momi, "Human-inspired active compliant and passive shared control framework for robotic contact-rich tasks in medical applications," *IEEE Transactions on Robotics*, 2025.

[24] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8971–8980, 2018.

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[26] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4282–4291, 2019.

[27] D. Raina, S. Chandrashekhara, R. Voyles, J. Wachs, and S. K. Saha, "Robotic sonographer: Autonomous robotic ultrasound using domain expertise in bayesian optimization," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6909–6915. IEEE, 2023.

[28] R. Martinez-Cantin, "Bayesian optimization with adaptive kernels for robot control," in *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 3350–3356. IEEE, 2017.

[29] J. Močkus, "On bayesian methods for seeking the extremum," in *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974 6*, pp. 400–404. Springer, 1975.

[30] J. Fu, M. Pecorella, E. Iovene, M. C. Palumbo, A. Rota, A. Redaelli, G. Ferrigno, and E. De Momi, "Augmented reality and human–robot collaboration framework for percutaneous nephrolithotomy: System design, implementation, and performance metrics," *IEEE Robotics & Automation Magazine*, 2024.

[31] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.

[32] W. Yan, Q. Ding, J. Chen, Y. Liu, and S. S. Cheng, "Needle tip tracking in 2d ultrasound based on improved compressive tracking and adaptive kalman filter," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3224–3231, 2021.

[33] W. Yan, Q. Ding, J. Chen, K. Yan, R. S.-Y. Tang, and S. S. Cheng, "Visual tracking of needle tip in 2d ultrasound based on global features in a siamese architecture," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4782–4788. IEEE, 2023.

[34] X. Duan, R. He, Q. Zhao, X. Chen, and C. Li, "Unified admittance control for accurate puncture and respiration following based on disturbance observation and model predictive control," *IEEE Robotics and Automation Letters*, 2025.

[35] S. Wilcox, Z. Huang, J. Shah, X. Yang, and Y. Chen, "Respiration-induced organ motion compensation: A review," *Annals of Biomedical Engineering*, vol. 53, no. 2, pp. 271–283, 2025.

[36] M. Khadem, C. Rossa, N. Usmani, R. S. Sloboda, and M. Tavakoli, "Semi-automated needle steering in biological tissue using an ultrasound-based deflection predictor," *Annals of Biomedical Engineering*, vol. 45, no. 4, pp. 924–938, 2017.

[37] M. Abayazid, P. Moreira, N. Shahriari, S. Patil, R. Alterovitz, and S. Misra, "Ultrasound-guided three-dimensional needle steering in biological tissue with curved surfaces," *Medical engineering & physics*, vol. 37, no. 1, pp. 145–150, 2015.