

Deep Kernel and Image Quality Estimators for Optimizing Robotic Ultrasound Controller using Bayesian Optimization

Deepak Raina^{1,2*}, SH Chandrashekhara³, Richard Voyles², Juan Wachs², Subir Kumar Saha¹

Abstract—Ultrasound is a commonly used medical imaging modality that requires expert sonographers to manually maneuver the ultrasound probe based on the acquired image. Autonomous Robotic Ultrasound (A-RUS) is an appealing alternative to this manual procedure in order to reduce sonographers’ workload. The key challenge to A-RUS is optimizing the ultrasound image quality for the region of interest across different patients. This requires knowledge of anatomy, recognition of error sources and precise probe position, orientation and pressure. Sample efficiency is important while optimizing these parameters associated with the robotized probe controller. Bayesian Optimization (BO), a sample-efficient optimization framework, has recently been applied to optimize the 2D motion of the probe. Nevertheless, further improvements are needed to improve the sample efficiency for high-dimensional control of the probe. We aim to overcome this problem by using a neural network to learn a low-dimensional kernel in BO, termed as Deep Kernel (DK). The neural network of DK is trained using probe and image data acquired during the procedure. The two image quality estimators are proposed that use a deep convolution neural network and provide real-time feedback to the BO. We validated our framework using these two feedback functions on three urinary bladder phantoms. We obtained over 50% increase in sample efficiency for 6D control of the robotized probe. Furthermore, our results indicate that this performance enhancement in BO is independent of the specific training dataset, demonstrating inter-patient adaptability.

I. INTRODUCTION

Ultrasound imaging delivers instantaneous, non-invasive access to the human body to effectively visualize anatomy and pathology [1]. Ultrasound stands out among other imaging modalities, such as MRI or CT scan, because it is low-cost, non-radiating, and suitable for use in rural or underserved communities. However, unlike other modalities, ultrasound is not a plug-and-play technology. Acquiring quality images relies heavily on the expertise of the trained sonographer who ensures optimal ultrasound probe contact location, pressure, angle, and gel condition between the probe and skin [2]. In addition, the sonographer must integrate the visual feedback from the image with a mental model of anatomical structure(s) to continuously improve the probing process.

In the pursuit of reducing the dependency on experts, enhancing access to care and eliminating the requirement for

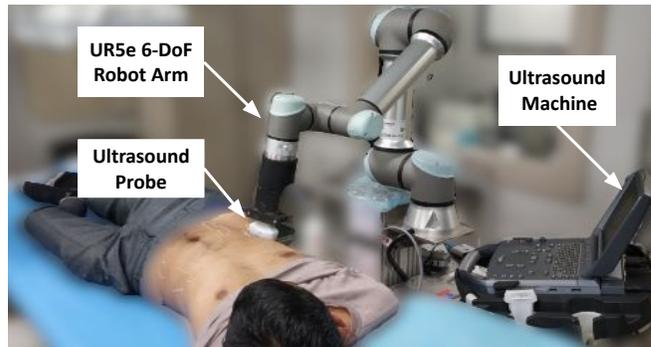


Fig. 1: Robotic ultrasound system with probe attached to its end-effector [2], conducting an *in-vivo* ultrasound.

direct patient contact, a Robotic Ultrasound System (RUS) is introduced [3]. RUS has a 6-DoF robotic arm with an ultrasound probe attached to its end-effector, as shown in Fig. 1. Recently, several telerobotic or human-assisted ultrasound systems have been proposed [2], [4]–[6]. In contrast to those systems, our system is fully autonomous. This allows for shorter procedural times, lower cognitive load, and prevents communication delays. However, several challenges need to be addressed for autonomous RUS. One of the key challenges is high-dimensional control of the probe maneuvers over the patient body. The probe position, angle and pressure need to be adjusted precisely and uniquely for inter- and intra-patient procedures. Inappropriate probe maneuvers would lead to acquiring low-quality images having artifacts, thereby leading to false alarms during the patients’ diagnosis [7]. In order to master this procedure’s skill, novice operators repeatedly scan different subjects in laboratory and clinical setting during their medical training [8]. Inspired by their training method, several model-based [9], [10] and model-free [11]–[13] paradigms have been applied to learn the robotized probe maneuvering for acquiring the optimal image. However, these techniques either provide limited probe motions or require a prohibitive number of samples for training, making them infeasible for in-human optimization of the RUS controller. This has led to the research community’s interest in sample-efficient learning techniques for optimizing the RUS controllers.

Bayesian Optimization (BO) is a popular sample-efficient, gradient-free and black-box optimization method for learning robotic controller parameters [14], [15]. Recently, BO has been applied to optimize the 2D motion of robotized ultrasound probe [16]. The clinical procedures, however, require high-dimensional motion of probe for adjusting its position, orientation and forces. Moreover, the performance

This work was supported in part by SERB (India) - OVDF Award No. SB/S9/Z-03/2017-VIII; PMRF - IIT Delhi under Ref. F.No.35-5/2017-TS.I:PMRF; National Science Foundation (NSF) USA under Grant #2140612; Daniel C. Lewis Professorship and PU-IUPUI Seed Grant.

¹Indian Institute of Technology (IIT), Delhi, India ({deepak.raina, saha}@mech.iitd.ac.in); ²Purdue University (PU), Indiana, USA ({draina, rvoyles, jpwachs}@purdue.edu); ³All India Institute of Medical Sciences (AIIMS), Delhi, India (drchandraradioaiims@gmail.com).

of BO degrades in high dimensions due to its exponentially increasing statistical and computational complexity [17], [18]. In this work, we formulate a transformation that uses a neural network to construct a low-dimensional kernel metric, thereby providing a sample-efficient BO framework for optimizing a high-dimensional RUS controller. The kernel metric is termed as Deep Kernel (DK), which is learned by utilizing the probe and image data obtained offline from the ultrasound procedure. This requires maneuvering the ultrasound probe over the scanning region and recording the probe poses and corresponding images. The neural network then learns a mapping between the high-dimensional probe pose and low-dimensional image quality metric. In order to estimate image quality, we propose two estimators using a Deep Convolution Neural Network (D-CNN) for image classification and segmentation, respectively, which also provide real-time feedback to BO. In order to validate it experimentally, we optimized a 6D control of the robotized probe, including position, orientation and force along z -axis, for three urinary bladder phantoms. We demonstrate that using a Deep Kernel in BO outperforms the sample-efficiency of standard kernel formulation. We noted that DK's performance is independent of a specific training dataset, which shows inter-patient adaptability. Finally, we validated the proposed image quality estimators on the phantom dataset and noted enhancement in accuracy over state-of-the-art networks.

A. Related Work

1) Autonomous Robotic Ultrasound (A-RUS) systems:

Several A-RUS systems have been developed to address the challenges of manual procedure. Recent works have used confidence-maps [9], [10], [19] or support vector machines [20] to assess image quality and then used it for adjusting the probe poses and forces. However, the probe control is limited, having either 2D/3D motion or decoupled poses/forces, which is insufficient to acquire a good quality ultrasound image for complex procedures on patients with challenging physiological conditions. Li *et al.* [11], [12] have used a deep reinforcement learning framework to develop a robotized probe controller for optimizing the image quality in spinal ultrasound. However, these systems require a large number of explorations to find the global optimum, therefore their success is limited to phantoms. Jiang *et al.* [21], [22] have used segmentation of tubular structures in human limbs for scanning them while the limb is static or moving. However, the procedures like spine or limb ultrasound require the probe to be oriented in a normal direction to the point of contact, with minimal orientation adjustment required throughout scanning. In contrast to these systems, we propose a generic A-RUS with a comprehensive probe controller using a sample-efficient BO framework.

2) *Bayesian optimization for ultrasound robots:* BO has been used for several safety-critical robotic medical procedures, such as autonomous robotic palpation [23], semi-autonomous surgical robot [24], controller tuning of hip exoskeletons [25] and A-RUS [16], [26]. Our work is a non-trivial extension to work by Goel *et al.* [16]. They

proposed using BO for A-RUS utilizing segmentation of the vessel in the ultrasound image as feedback to the BO for finding and scanning the region with high vessel density. They used hybrid position-force control to move the robot in 2D plane while maintaining constant force along the probe axis normal to the point of contact. In contrast, our work proposes technical enhancements to make this approach more practical and expand its scope to complex ultrasound procedures. We propose a 6D control of the probe pose, including the variable force control along the z -axis. We further propose a novel kernel for BO to improve its sample efficiency during the optimization of 6D probe controller.

3) *High-Dimensional Bayesian Optimization:* A large body of literature has addressed the issues related to the scalability of BO to high-dimensional optimization problems. A few methods have exploited the potential additive structure in the objective function [27], [28]. However, these methods require partitioning of the objective function and a large number of GPs, hence they do not generalize well to several applications. The other methods determined the mapping from high-dimensional space to low-dimensional subspace [29], [30]. For robotic systems, the transformation of search space to a lower dimension has been attempted for locomotion controller [31], [32]. Rai *et al.* [32] proposed a transformation based on human walking knowledge to project a 16D controller to 1D controller. Antonova *et al.* [31] propose to learn the transformation using a neural network trained on data obtained from a high-fidelity simulator. However, the paradigms of high-dimensional BO have not yet been explored for optimizing the RUS controller.

II. METHODOLOGY

The overview of the proposed BO is shown in Fig. 2. In the *offline phase*, first the dataset of images will be collected from the scanning region by varying the probe poses. Then, each image in dataset will be annotated for quality rating and segmentation masks, which will be used to train the *Image Quality Estimators (IQEs)*. Second, the *Deep Kernel (DK)* neural network is trained using the dataset of probe poses and their corresponding image qualities estimated using the IQEs trained earlier. In the *online phase*, the BO will utilize the DK in Gaussian Process (GP) model and learnt IQEs to calculate the probabilistic estimate of the unknown image quality in the scanning region of human body. An *acquisition function* is optimized during online BO to yield the new probing pose, which will acquire the new ultrasound image. The GP model is then re-fitted to the new data and process is repeated until the termination criteria is reached, which can either be the maximum quality reached or maximum reasonable steps.

A. Bayesian optimization

The BO can determine the probing poses that will optimize the ultrasound image quality within a scanning region. If A denotes the scanning region on the human body, then the objective function of BO is given by:

$$\max_{\mathbf{p} \in A} q(\mathcal{I}(\mathbf{p})) \quad (1)$$

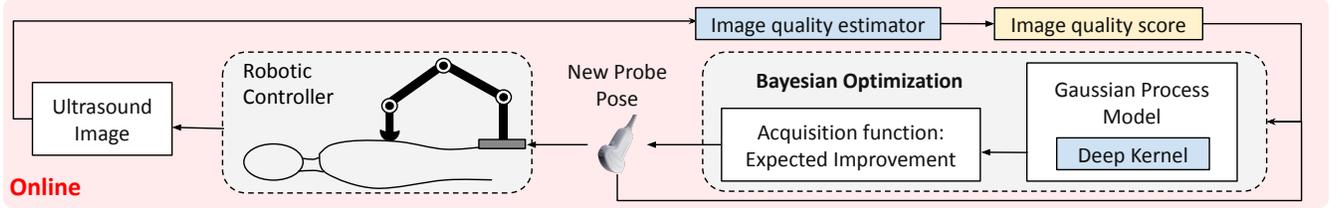


Fig. 2: Overview of the Bayesian optimization (BO) framework for optimizing 6D robotic ultrasound controller using deep kernels in Gaussian process model and image quality estimators as feedback.

where $q(\mathcal{I}(\mathbf{p}))$ denotes the ultrasound image quality score \mathcal{I} at probing pose \mathbf{p} .

1) *Deep Kernel*: The GP model is used as the function estimator in BO. It defines an unknown quality function value $q(\mathcal{I}(\mathbf{p}))$ at a probe pose \mathbf{p} by a mean function $\mu(\cdot)$ and kernel function $\kappa(\cdot, \cdot)$ as

$$q(\mathcal{I}(\mathbf{p})) \sim GP(\mu(\cdot), \kappa(\cdot, \cdot)) \quad (2)$$

Given the quality values set $\bar{q} = [q(\mathcal{I}(\mathbf{p}_1)), \dots, q(\mathcal{I}(\mathbf{p}_n))]$ at probe poses $\bar{\mathbf{p}} = [\mathbf{p}_1, \dots, \mathbf{p}_n]$, the GP regression can predict the quality function value at new probe pose \mathbf{p}^* as the Gaussian distribution and is given by:

$$\mathcal{P}(q(\mathcal{I}(\mathbf{p}^*)) | \mathbf{p}^*, \bar{\mathbf{p}}, \bar{q}) = \mathcal{N}(\mathbf{k}\mathbf{K}^{-1}\bar{q}, \kappa(\mathbf{p}^*, \mathbf{p}^*) - \mathbf{k}\mathbf{K}^{-1}\mathbf{k}^T) \quad (3)$$

where,

$$\mathbf{k} = [\kappa(\mathbf{p}_*, \mathbf{p}_1) \quad \dots \quad \kappa(\mathbf{p}_*, \mathbf{p}_n)]$$

$$\mathbf{K} = \begin{bmatrix} \kappa(\mathbf{p}_1, \mathbf{p}_1) & \dots & \kappa(\mathbf{p}_1, \mathbf{p}_n) \\ \vdots & \ddots & \vdots \\ \kappa(\mathbf{p}_n, \mathbf{p}_1) & \dots & \kappa(\mathbf{p}_n, \mathbf{p}_n) \end{bmatrix}$$

For kernel matrix κ , we propose using a sum of two functions, the radial basis function and white noise function, because their combination provides better estimates for anatomical structures in ultrasound images [16].

$$\kappa(\mathbf{p}_i, \mathbf{p}_j) = \sigma_r \exp\left(\frac{-\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2l^2}\right) + \sigma_w \mathbf{I}_n \quad (4)$$

where σ_r , σ_w and l is the overall variance, noise variance and length-scale, respectively, representing the BO hyperparameters θ . Robotized control of ultrasound probe necessitates high-dimensional control of probe pose, including its position, orientation and forces to acquire appropriate quality images for inter- and intra-patient procedures. However, the performance of BO often degrades in high dimensions [18], specifically due to two reasons: (i) the high dimensional objective function exponentially increases the number of samples (queries) required to explore the space (ii) the BO's acquisition function is non-convex and optimizing it demands exponentially increasing computational power.

In order to overcome these issues, we propose a 1D kernel for GP and term it as a Deep Kernel, as shown in Fig. 3. It will transform the 6D representation of probe poses into a 1D space. We use a supervised Neural Network (NN), represented by ϕ_D , which will learn the mapping between probe pose representation and image quality metric. The DK

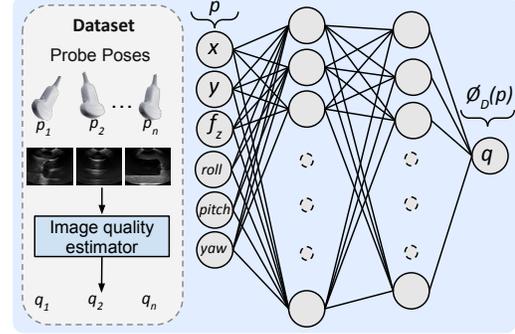


Fig. 3: Dataset and neural network structure for Deep Kernel

neural network used two hidden layers, each having 400 and 300 neurons, respectively. The activation function used for each layer is Rectified Linear Unit (ReLU). The estimation of image quality metric for a given image has been explained in section II-B. The data for training the neural network can be collected offline by the expert-demonstrated probing maneuvers or random sampling of the probe poses over the scanning region. The functional form of this kernel is identical to that of eq. (4). However, the euclidean difference between the six-dimensional probe poses was replaced with the difference between the uni-dimensional NN output image quality score for the input probe pose.

$$\kappa(\mathbf{p}_i, \mathbf{p}_j) = \sigma_r \exp\left(\frac{-\|\phi_D(\mathbf{p}_i) - \phi_D(\mathbf{p}_j)\|^2}{2l^2}\right) + \sigma_w \mathbf{I} \quad (5)$$

where, $\phi_D(\mathbf{p}_*)$ denotes the output of NN at probe pose \mathbf{p}_* . The hyper-parameters θ of DK-based GP will be estimated by maximizing the log marginal likelihood with a Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) solver, which is given by

$$\theta^* = \arg \max_{\theta} \log \prod \mathcal{N}(q(\mathcal{I}(\mathbf{p}_i)) | \mu_{\theta}(\mathbf{p}_i), \mathbf{K}) \quad (6)$$

2) *Acquisition Function*: The next probe pose is determined by optimizing the acquisition function. We have used an Expected Improvement (EI), which is the most commonly used acquisition function. Given the posterior mean and variance of GP as $\mu_{\bar{q}}(\mathbf{p})$, $\sigma_{\bar{q}}^2(\mathbf{p})$, the expression for EI is:

$$EI(\mathbf{p}) = \begin{cases} (\mu_{\bar{q}}(\mathbf{p}) - q^+ - \xi)\Phi(\mathbf{Z}) + \sigma_{\bar{q}}^2(\mathbf{p})\phi(\mathbf{Z}) & \text{if } \sigma_{\bar{q}}^2(\mathbf{p}) > 0 \\ 0 & \text{if } \sigma_{\bar{q}}^2(\mathbf{p}) = 0 \end{cases} \quad (7)$$

where $\mathbf{Z} = \frac{\mu_{\bar{q}}(\mathbf{p}) - q^+}{\sigma_{\bar{q}}^2(\mathbf{p})}$ if $\sigma_{\bar{q}}^2(\mathbf{p}) > 0$ else 0; Φ and ϕ are the Probability Density Function (PDF) and Cumulative Density Function (CDF) of standard normal distribution, respectively and $q^+ = q^+(\mathcal{I}(\mathbf{p}))$ is the best observed image quality so

far. The exploration and exploitation during optimization is balanced by parameter ξ , and high ξ means more exploration or less exploitation.

B. Image quality estimation:

We proposed two image quality estimators for estimating real-time diagnostic ultrasound image quality in A-RUS, which are based on: (1) Classification and (2) Segmentation of ultrasound image, as shown in Fig. 4.

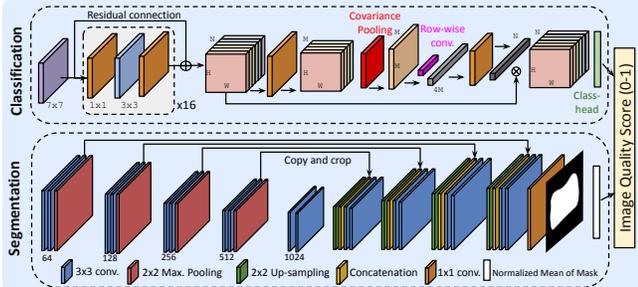


Fig. 4: D-CNNs for image quality estimators

1) *Image classification*: We propose using a Deep Convolutional Neural Network (D-CNN) based classifier to extract the ultrasound image features and classify them based on the diagnostic quality. Ultrasound image quality assessment requires rich feature extraction for classifying the images that look quite similar but differ a lot in terms of image quality. In recent work, Song *et al.* [33] proposed a Bilinear Convolutional Neural Network (B-CNN) for fine-grained classification of breast ultrasound image quality. We propose a statistical enhancement to this work to improve its performance for ultrasound images having complex anatomical structures. Recently, Li *et al.* [34] proposed Second-order Pooling (SoP) block for enhancing the non-linear modeling ability of D-CNNs. Inspired by their work, we perform the SoP along spatial dimension at the end of the D-CNN network. First, we reduce the number of channels of feature volume $\mathbf{X} \in \mathbb{R}^{H \times W \times N}$ from N to M using 1×1 convolution. Then, \mathbf{X} is reshaped to $\mathbf{X} \in \mathbb{R}^{M \times N}$ where $M = H \times W$. Later, we compute the covariance matrix $\mathbf{C}_{M \times M} = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T$, reshaped it to $1 \times M \times M$ and passed it through convolution layer with $4M$ kernels of size $1 \times M$. The resulting tensor $1 \times 1 \times 4M$ is then resized to $1 \times 1 \times M$ tensor, \mathbf{w}_M , using a 1×1 convolution. Finally, weight vector \mathbf{w}_M is multiplied with \mathbf{X} to obtain the resultant feature volume $\mathbf{V}_M \in \mathbb{R}^{H \times W \times N}$. The base network used in proposed D-CNN is Residual Network (ResNet50). We trained the network using the Categorical Cross Entropy as a loss function.

2) *Image segmentation*: To evaluate the image quality based on image segmentation, we used U-net [35], which is a state-of-the-art segmentation network for biomedical image segmentation. The network consists of two 2D convolution layer (Conv2D), followed by a Batch Normalization layer, Rectified Linear Unit (ReLU) as activation function, a max. pooling layer (i.e down sampling) of 2×2 in the contracting path. Each Conv2D layer is padded convolution with kernel size as 3×3 . Similarly, for the expansive path, an up-sampling layer of 2×2 was added after each Conv2D layer.

The number of filters along the contracting path doubled with each successive convolutional layer. The inverse of this pattern was replicated in the expanding path, resulting in the same number of filters in the last 3×3 convolutional layer of expanding path. In our dataset images, the bladder boundaries are blurred and irregular, making its segmentation rather challenging. Hence, we have used multiple loss functions to train the network, so that the model would be enforced to learn fine-grained features of the bladder [36]. We used the combinations of three loss functions, Dice coefficient (DC) and Jaccard Index (JI) and Binary Cross-Entropy (BCE), formulated as $L_{DJB} = L_{DC} + L_{JI} + L_{BCE}$, where

$$L_{DC} = 1 - \left(\frac{2y_t y_p + s}{y_t + y_p + s} \right), L_{JI} = 1 - \frac{y_t y_p + s}{y_t + y_p - y_t y_p + s}$$

$$L_{BCE} = -y_t \log(y_p) - (1 - y_t) \log(1 - y_p)$$

where s , y_t and y_p represent smoothness constant, ground truth and predicted mask, respectively. The model will finally output the predicted mask of the given input ultrasound image. We take normalized mean of the predicted mask as the quality score, the value of which lies between 0 and 1.

3) *Dataset*: We collected ultrasound image dataset from Urinary Bladder (UB) phantom (YourDesignMedical, USA). A total of 2290 phantom images were collected. All images have been resized to 224×224 shape. The training and validation set consists of 2061 and 229 images, respectively. For classification, the ground truth quality of images is an integer score between 1 – 5, based on an internationally prescribed generalized 5-level absolute assessment scale [4]. The training and validation set contains an equal distribution of images for each quality type. A score of 1 means no appearance of the UB with unacceptable artifacts and 5 means that the clear depiction of the UB with distinct boundaries and acceptable artifacts, depicting a high diagnostic accuracy. A poor-quality image either contains noise or motion artifacts, blurred images, indistinct boundaries, obscuring the posterior or anterior sections of the UB. Later, we normalized the quality score in the range 0 – 1 for comparison with segmentation-based quality score, where 0 denotes a score when the force value of probe along z-axis is below minimum required for appropriate contact. For segmentation, we annotated using SuperAnnotate (<https://superannotate.com/>). A single polygon with multiple points is drawn in every image, spanning the shape of UB. These polygons were then used to generate corresponding ground truth masks.

C. Robotic controller

The robotic controller will move the probe to the new pose $\mathbf{p} = [x, y, z, roll, pitch, yaw]$ given by BO. The hybrid position-force control is used for controlling the robot, where only z -axis is under force control, denoted as f_z . For the safety of phantoms, the force limits have been set to $20N$.

III. RESULTS AND DISCUSSIONS

A. Experimental setup

We validate our framework on an experimental setup of robotic ultrasound system, consisting of a 7-DoF Sawyer

collaborative robotic arm (Rethink Robotics, Germany) with a Micro Convex MC10-5R10S-3 probe attached to its end-effector. The probe is connected to the Telemed Ultrasound machine (Telemed Medical Systems, Italy), which captures the ultrasound image. The machine is connected to the laptop for displaying and transmitting the image for processing in BO. The ultrasound scanning is conducted on a urinary bladder phantom. In order to validate the proposed framework for

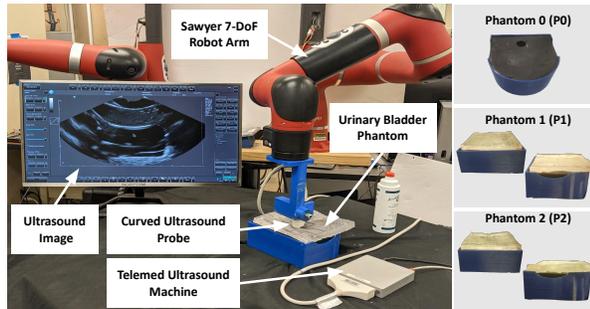


Fig. 5: Experimental setup of robotic ultrasound system with three different phantoms of the urinary bladder.

inter-patient ultrasound scanning, we modified the phantom $P0$ using the layers of ballistic gel. We used 0.38 inch layer for $P1$ and 0.63 inch layer for $P2$, as shown in Fig. 5. The BO has been implemented in Python 3.8. The DK neural network and IQEs explained in Section II has been trained using PyTorch 1.11. For BO, we used $\xi = 0.1$ and region limits A as $x \in (-0.05, 0.05)m$, $y \in (-0.02, 0.02)m$, $f_z \in (5 - 20)N$, $roll \in (-0.2, 0.2)rad$, $pitch \in (-0.2, 0.2)rad$, and $yaw \in (-0.5, 0.5)rad$. The orientation limits have been decided to ensure collision-free scanning of phantoms.

B. Analyzing BO performance with Deep kernel

In order to analyze the effectiveness of the proposed framework, we compared the performance of BO that used our proposed neural network based Deep kernel with a standard Radial Basis Function (RBF) kernel. We have analysed our results for two feedback functions, as described in Section II-B: (i) Image quality score obtained from classification network (q_c), which is a non-smooth feedback; (ii) Image segmentation score obtained from segmentation network (q_s), which is a smooth feedback function. The dataset for deep kernel has been sampled from phantoms $P0$ and $P1$, while no dataset has been taken from $P2$, which is our test-case phantom. For training the deep kernel, we sampled 1200 probe poses, 600 for each of the phantom $P0$ and $P1$,

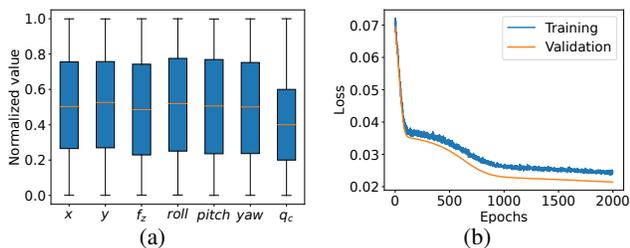


Fig. 6: (a) Distribution of probe poses sampled for training; (b) learning curve of deep kernel's neural network

using latin hypercube sampling and scanned the phantoms using sampled probe poses. The distribution of probe-poses sampled for each pose variable and corresponding quality is shown in Fig. 6a. The distribution shows that probe control parameters have been widely explored. However, out of the 1200 sampled poses, only 110 resulted in a quality greater than 0.8. This means that random sampling has only 9.2% chance of locating a high image quality region. This shows that acquiring high-quality images requires all sets of probe parameters to be accurate, which highlights the complexity of this optimization problem. The learning curve for the deep kernel's neural network is shown in Fig. 6b.

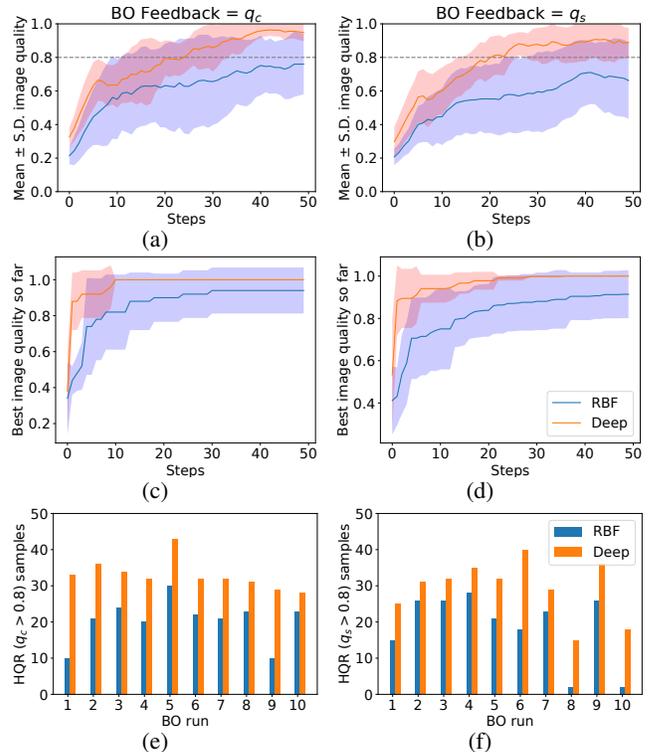


Fig. 7: Analysis of BO runs on phantom $P0$ and $P1$ using classification q_c (a,c,e) and segmentation q_s (b,d,f) feedback. Line show mean and shaded region shows \pm S.D over 10 runs

In total, we conducted 60 runs of BO, 5 each for 3 phantoms and 2 feedback functions. Each run of BO has 50 steps, which is the termination condition. In this section, we denote the quality score greater than 0.8 as *high quality region (HQR)*. The average image quality achieved during the runs of BO for phantom $P0$, $P1$ are shown in Fig 7a and 7b. It indicates that BO with RBF kernel maintains a low value of average image quality compared to deep kernel over all steps of BO, and hence it could not locate the HQR reliably even after 50 steps. Further, the large values of standard deviation (represented by the shaded region on both sides of the mean line) for RBF kernel show its extensive exploratory nature. In contrast, deep kernel maintains high mean quality with low standard deviations, which shows its better sample efficiency. The deep kernel also succeeded in locating the HQR in first 5 – 10 steps in all runs of BO, as shown in Fig. 7c and 7d. Further, Fig. 7e and 7f shows that deep kernel has 61.76%

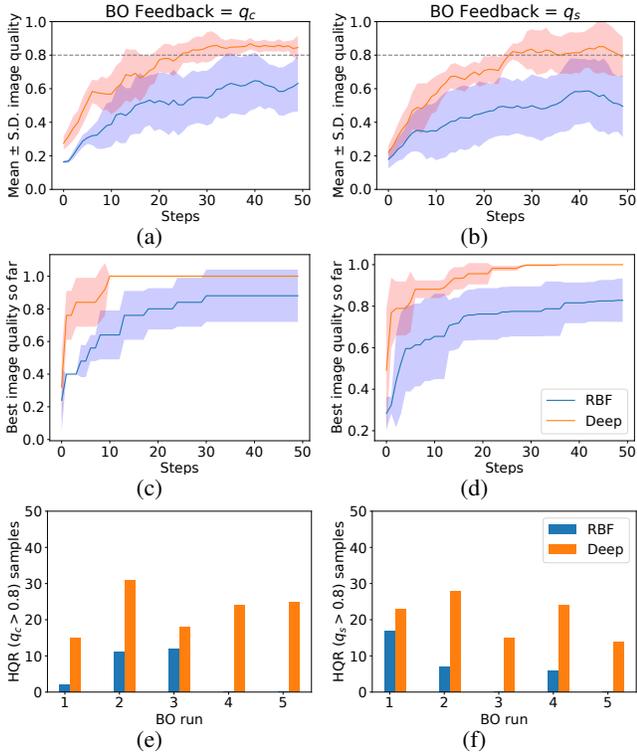


Fig. 8: Analysis of BO runs on phantom $P2$ using classification q_c (a,c,e) and segmentation q_s (b,d,f) as feedback. Line show mean and shaded region shows \pm S.D over 10 runs

and 56.68% more sample in HQR than RBF for q_c and q_s feedback, respectively, in 10 runs of BO.

For phantom $P2$, the average values in Fig. 8a and 8b show that BO with RBF kernel could not locate the HQR reliably (high S.D.). However, BO with deep kernel reached the HQR in approximately 10 – 30 steps, as shown in Fig. 8c and 8d. In contrast, the RBF kernel took more than 30 steps to locate the HQR reliably. Moreover, out of the 5 runs, RBF kernel could only locate the HQR in 3/5 runs for q_c and q_s , as shown in Fig. 8e and Fig. 8f. Further, RBF has less number of HQR samples in comparison to deep kernel across all runs of BO. These results indicate that the deep kernel improves the sample efficiency of BO and its performance is independent of its training dataset.

C. Analysis of image quality estimators

We evaluated the classification network on a 10-fold cross-validation for classifying the urinary bladder phantom images based on their quality. To avoid over-fitting on the small dataset, we applied random horizontal flipping and normalization data augmentations. The values of precision, recall, and accuracy for five classes of quality are shown in Table I. We compared the proposed network, ResNet50 with Spatial SoP (RN50+SSoP), with the ResNet50 with Bilinear Pooling (RN50+BP) [33], to validate the performance enhancement. Results show that the RN50+SSoP has an average accuracy of 98.2 ± 0.5 , which is 1.42% more than the average accuracy of RN50+BP. The class-wise analysis suggest that RN50+SSoP has significant gains over the RN50+BP, achieving 1.46%, 2.06%, 2.95% and 0.91%

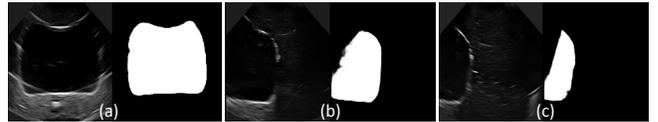


Fig. 9: The predicted masks when bladder appears (a) at full size, (b) approx. half of its size, and (c) one-fourth of its size

more accuracy for quality scores 2, 3, 4 and 5, respectively. Hence, it is validated that the proposed network enhanced the accuracy of ultrasound image quality classification.

TABLE I: Comparison of the classification-based IQE with the state-of-the art network on 10-fold cross validation

Score	RN50+BP [33]			RN50+SSoP (Proposed)		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
1	99.7 ± 0.3	99.4 ± 0.7	99.5 ± 0.4	99.7 ± 0.5	99.9 ± 0.3	99.8 ± 0.4
2	93.9 ± 2.1	95.8 ± 1.7	94.8 ± 1.3	95.5 ± 1.7	97.0 ± 1.5	96.2 ± 1.3
3	94.6 ± 1.3	95.5 ± 1.9	95.0 ± 1.1	97.8 ± 1.3	96.3 ± 1.6	97.0 ± 1.2
4	96.3 ± 1.4	94.8 ± 1.6	95.5 ± 0.7	98.7 ± 0.9	98.1 ± 0.4	98.4 ± 0.3
5	97.3 ± 0.8	98.9 ± 0.3	98.0 ± 0.1	98.7 ± 0.6	99.1 ± 0.2	98.9 ± 0.1
Avg.	96.5 ± 0.5	97.1 ± 0.5	96.8 ± 0.5	98.2 ± 0.5	98.2 ± 0.5	98.2 ± 0.5

For the segmentation network, the qualitative results are shown in Fig. 9. We further evaluated this network quantitatively using the standard evaluation matrices, including Precision (P), Recall (R), F1 score (F1), Dice Coefficient (DC), and Intersection over Union (IoU). The values of these evaluation metrics for the 10-fold cross-validation and comparison with standard dice loss function (L_{DC}) are given in Table II. The proposed loss function, L_{DJB} outperforms L_{DC} on P, R, F1, DC and IoU by 3.2%, 4.7%, 2.6%, 2.5% and 4.8%, respectively. Hence, the U-net guided with the proposed multi-loss function is an effective method for the segmentation of complex structures in the ultrasound image.

TABLE II: Comparison of the segmentation-based IQE for standard dice loss and proposed multi-loss function

Loss	P	R	F1	DC	IoU
L_{DC}	95.6 ± 6.09	92.7 ± 5.55	93.9 ± 2.99	93.7 ± 3.04	88.4 ± 5.41
L_{DJB}	98.8 ± 0.26	97.3 ± 2.02	96.4 ± 1.57	96.1 ± 2.10	92.9 ± 1.81

IV. CONCLUSION

We proposed a sample-efficient Bayesian Optimization (BO) framework using deep kernel and image quality estimators for optimizing the 6-dimensional robotic ultrasound controller. The deep kernel of Gaussian process model learns a one-dimensional embedding representing image quality for a given probe pose. The two image quality estimators (IQEs) based on classification and segmentation utilizing a deep convolutional neural network are proposed for real-time quality assessment in RUS. We validated the framework on three urinary bladder phantoms and demonstrated over 50% increase in sample efficiency over the standard kernel for both IQEs as feedback to BO. In future work, we would like to validate it for complicated surface phantoms and *in-vivo* trials for a clinical procedure [2], [37]. We will explore approaches for improving the IQEs and learning the kernel from multiple expert demonstrations. We would also investigate integrating the deep kernel with a pre-trained prior quality model similar to [38] for further improvement in the sample efficiency of BO.

REFERENCES

- [1] V. Chan and A. Perlas, "Basics of ultrasound imaging," in *Atlas of ultrasound-guided procedures in interventional pain management*, pp. 13–19, Springer, 2011.
- [2] D. Raina, H. Singh, S. K. Saha, C. Arora, A. Agarwal, S. Chandrashekhara, K. Rangarajan, and S. Nandi, "Comprehensive telerobotic ultrasound system for abdominal imaging: Development and in-vivo feasibility study," in *2021 International Symposium on Medical Robotics (ISMR)*, pp. 1–7, IEEE, 2021.
- [3] M. C. Roshan, A. Pranata, and M. Isaksson, "Robotic ultrasonography for autonomous non-invasive diagnosis—a systematic literature review," *IEEE Transactions on Medical Robotics and Bionics*, 2022.
- [4] S. Duan, L. Liu, Y. Chen, L. Yang, Y. Zhang, S. Wang, L. Hao, and L. Zhang, "A 5g-powered robot-assisted teleultrasound diagnostic system in an intensive care unit," *Critical Care*, vol. 25, no. 1, pp. 1–9, 2021.
- [5] J. Carriere, J. Fong, T. Meyer, R. Sloboda, S. Husain, N. Usmani, and M. Tavakoli, "An admittance-controlled robotic assistant for semi-autonomous breast ultrasound scanning," in *2019 international symposium on medical robotics (ISMR)*, pp. 1–7, IEEE, 2019.
- [6] T. Li, X. Meng, and M. Tavakoli, "Dual mode p hri-tele hri control system with a hybrid admittance-force controller for ultrasound imaging," *Sensors*, vol. 22, no. 11, p. 4025, 2022.
- [7] A. Pinto, F. Pinto, A. Faggian, G. Rubini, F. Caranci, L. Macarini, E. A. Genovese, and L. Brunese, "Sources of error in emergency ultrasonography," *Critical ultrasound journal*, vol. 5, no. 1, pp. 1–5, 2013.
- [8] U. Tarique, B. Tang, M. Singh, K. M. Kulasegaram, and J. Ailon, "Ultrasound curricula in undergraduate medical education: a scoping review," *Journal of Ultrasound in Medicine*, vol. 37, no. 1, pp. 69–82, 2018.
- [9] P. Chatelain, A. Krupa, and N. Navab, "Confidence-driven control of an ultrasound probe," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1410–1424, 2017.
- [10] Z. Jiang, M. Grimm, M. Zhou, J. Esteban, W. Simson, G. Zahnd, and N. Navab, "Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1342–1349, 2020.
- [11] K. Li, J. Wang, Y. Xu, H. Qin, D. Liu, L. Liu, and M. Q.-H. Meng, "Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8302–8308, IEEE, 2021.
- [12] K. Li, Y. Xu, J. Wang, D. Ni, L. Liu, and M. Q.-H. Meng, "Image-guided navigation of a robotic ultrasound probe for autonomous spinal sonography using a shadow-aware dual-agent framework," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 1, pp. 130–144, 2021.
- [13] G. Ning, X. Zhang, and H. Liao, "Autonomic robotic ultrasound imaging system based on reinforcement learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2787–2797, 2021.
- [14] L. Roveda, M. Magni, M. Cantoni, D. Piga, and G. Bucca, "Human-robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via bayesian optimization," *Robotics and Autonomous Systems*, vol. 136, p. 103711, 2021.
- [15] F. Zahedi, D. Chang, and H. Lee, "User-adaptive variable damping control using bayesian optimization to enhance physical human-robot interaction," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2724–2731, 2022.
- [16] R. Goel, F. Abhimanyu, K. Patel, J. Galeotti, and H. Choset, "Autonomous ultrasound scanning using bayesian optimization and hybrid force control," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 8396–8402, IEEE, 2022.
- [17] B. Choffin and N. Ueda, "Scaling bayesian optimization up to higher dimensions: a review and comparison of recent algorithms," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2018.
- [18] M. Malu, G. Dasarathy, and A. Spanias, "Bayesian optimization in high-dimensional spaces: A brief survey," in *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pp. 1–8, IEEE, 2021.
- [19] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," *Medical image analysis*, vol. 16, no. 6, pp. 1101–1112, 2012.
- [20] M. Akbari, J. Carriere, T. Meyer, R. Sloboda, S. Husain, N. Usmani, and M. Tavakoli, "Robotic ultrasound scanning with real-time image-based force adjustment: quick response for enabling physical distancing during the covid-19 pandemic," *Frontiers in Robotics and AI*, vol. 8, p. 645424, 2021.
- [21] Z. Jiang, Z. Li, M. Grimm, M. Zhou, M. Esposito, W. Wein, W. Stechele, T. Wendler, and N. Navab, "Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 7, pp. 7064–7075, 2021.
- [22] Z. Jiang, Y. Gao, L. Xie, and N. Navab, "Towards autonomous atlas-based ultrasound acquisitions in presence of articulated motion," *IEEE Robotics and Automation Letters*, 2022.
- [23] Y. Yan and J. Pan, "Fast localization and segmentation of tissue abnormalities by autonomous robotic palpation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1707–1714, 2021.
- [24] J. Chen, D. Zhang, A. Munawar, R. Zhu, B. Lo, G. S. Fischer, and G.-Z. Yang, "Supervised semi-autonomous control for surgical robot based on banoian optimization," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2943–2949, IEEE, 2020.
- [25] Y. Ding, M. Kim, S. Kuindersma, and C. J. Walsh, "Human-in-the-loop optimization of hip assistance with a soft exosuit during walking," *Science robotics*, vol. 3, no. 15, p. eaar5438, 2018.
- [26] Y. Huang, W. Xiao, C. Wang, H. Liu, R. Huang, and Z. Sun, "Towards fully autonomous ultrasound scanning robot with imitation learning based on clinical protocols," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3671–3678, 2021.
- [27] K. Kandasamy, J. Schneider, and B. Póczos, "High dimensional bayesian optimisation and bandits via additive models," in *International conference on machine learning*, pp. 295–304, PMLR, 2015.
- [28] Z. Wang, C. Gehring, P. Kohli, and S. Jegelka, "Batched large-scale bayesian optimization in high-dimensional spaces," in *International Conference on Artificial Intelligence and Statistics*, pp. 745–754, PMLR, 2018.
- [29] Z. Wang, F. Hutter, M. Zoghi, D. Matheson, and N. De Freitas, "Bayesian optimization in a billion dimensions via random embeddings," *Journal of Artificial Intelligence Research*, vol. 55, pp. 361–387, 2016.
- [30] A. Nayebi, A. Munteanu, and M. Poloczek, "A framework for bayesian optimization in embedded subspaces," in *International Conference on Machine Learning*, pp. 4752–4761, PMLR, 2019.
- [31] R. Antonova, A. Rai, and C. G. Atkeson, "Deep kernels for optimizing locomotion controllers," in *Conference on Robot Learning*, pp. 47–56, PMLR, 2017.
- [32] A. Rai, R. Antonova, S. Song, W. Martin, H. Geyer, and C. Atkeson, "Bayesian optimization using domain knowledge on the atrias biped," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1771–1778, IEEE, 2018.
- [33] Y. Song, Z. Zhong, B. Zhao, P. Zhang, Q. Wang, Z. Wang, L. Yao, F. Lv, and Y. Hu, "Medical ultrasound image quality assessment for autonomous robotic screening," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6290–6296, 2022.
- [34] Z. Gao, J. Xie, Q. Wang, and P. Li, "Global second-order pooling convolutional networks," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 3024–3033, 2019.
- [35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [36] D. Raina, K. Verma, S. H. Chandrashekhara, and S. K. Saha, "Slim u-net: Efficient anatomical feature preserving u-net architecture for ultrasound image segmentation," *arXiv preprint arXiv:2302.11524*, 2023.
- [37] S. H. Chandrashekhara, K. Rangarajan, A. Agrawal, S. Thulkar, S. Gamanagatti, D. Raina, S. K. Saha, and C. Arora, "Robotic ultrasound: An initial feasibility study," *World Journal of Methodology*, vol. 12, no. 4, pp. 274–284, 2022.
- [38] Y. Zhu, A. Smith, and K. Hauser, "Automated heart and lung auscultation in robotic physical examinations," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4204–4211, 2022.