2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Robotic Sonographer: Autonomous Robotic Ultrasound using Domain Expertise in Bayesian Optimization

Deepak Raina[12*], SH Chandrashekhara[3], Richard Voyles[2], Juan Wachs[2], Subir Kumar Saha[1]

***Abstract*— Ultrasound is a vital imaging modality utilized for a variety of diagnostic and interventional procedures. However, an expert sonographer is required to make accurate maneuvers of the probe over the human body while making sense of the ultrasound images for diagnostic purposes. This procedure requires a substantial amount of training and up to a few years of experience. In this paper, we propose an autonomous robotic ultrasound system that uses Bayesian Optimization (BO) in combination with the domain expertise to predict and effectively scan the regions where diagnostic quality ultrasound images can be acquired. The quality map, which is a distribution of image quality in a scanning region, is estimated using Gaussian process in BO. This relies on a prior quality map modeled using expert's demonstration of the high-quality probing maneuvers. The ultrasound image quality feedback is provided to BO, which is estimated using a deep convolution neural network model. This model was previously trained on database of images labelled for diagnostic quality by expert radiologists. Experiments on three different urinary bladder phantoms validated that the proposed autonomous ultrasound system can acquire ultrasound images for diagnostic purposes with a probing position and force accuracy of 98.7% and 97.8%, respectively.**

I. Introduction

Ultrasound is the most frequently used imaging modality for diagnostic and surgical interventions due to its low cost, non-ionizing nature, portability and real-time feedback. Ultrasound offers several advantages over other imaging modalities, like Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), however, the diagnosis by ultrasound is a highly operator-dependent modality [1]. This is because of the skills required for manual control of the probe and quality assessment of acquired images. Sonographers employ both directed as well as random explorations strategies to search for diagnostic-quality images. The ultrasound probe is moved within the region of interest through hand maneuvers initially and fine adjustments to the probe's translational and rotational motion later. These maneuvers also include the safe and precise adjustment of the pressure through the probe while simultaneously analyzing the quality of acquired images. Such an intricate procedure requires a great deal of skill, focus, experience and manual effort from sonographers. In rural settings, skilled sonographers availability is limited [2], and alternative solutions are required.

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.

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

*Corresponding author is Deepak Raina

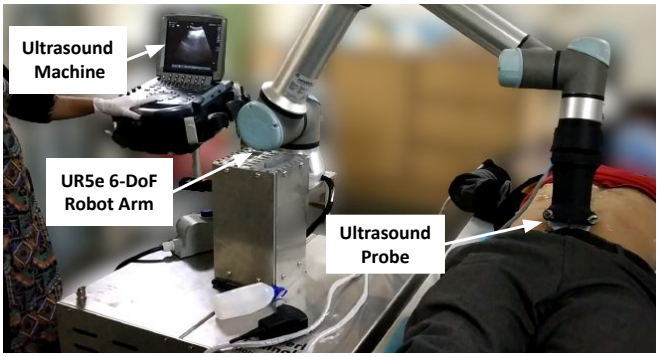

Fig. 1: Robotic ultrasound system with probe attached to its end-effector [3], conducting a urinary bladder ultrasound.

In order to reduce the burden on experts, a Robotic Ultrasound System (RUS) is introduced. RUS consists of a dexterous robotic arm and an ultrasound machine with its probe attached to the end effector of the robot, as shown in Fig. 1. RUS can help ensure the accuracy, safety and consistency of the ultrasound procedures. Recently, in order to address the aforementioned needs, several telerobotic or human-assisted ultrasound systems have been proposed [3]–[7]. Compared to these systems, a fully automated ultrasound system offers various potential benefits, including shorter procedure time, a shorter learning curve, minimal communication delays and a reduced cognitive load [8]. However, there are key challenges for effective autonomous RUS. One of the most important challenge has to do with the hand motions for ultrasound images acquisition. Such images exhibit considerable inter- and intra-subject variability and the image quality is highly dependent on the precise position, orientation and pressure of the ultrasound probe. With incorrect probe maneuvers, the resulting image presents noise, artifacts, blurred boundaries and poor visibility, thereby making it unacceptable for diagnosis. Sonographers rely on visual and haptic feedback, anatomical information, and diagnostic expertise from prior medical education to rapidly acquire the high-quality images. Therefore, the RUS must locate the regions with acceptable diagnostic image quality for inter- and intra-patient procedures in the fewest exploration steps.

In this paper, we present an autonomous robotic ultrasound system that uses the domain-expertise in Bayesian Optimization (BO)-based search to scan the anatomical regions for acquiring diagnostic quality ultrasound images, thereby eliminating the need to thoroughly scan the entire region. The *key contributions* of our work are as follows:

1) We proposed a prior in BO, gleaned from the expert's demonstration of high image quality probing poses,

termed as *expert's prior*. BO then estimates the region's unknown image quality as a semi-parametric Gaussian process model with expert's prior.
2) A novel *image quality metric* is proposed, trained using a dataset of ultrasound images labelled for diagnostic quality by expert radiologists, which provides image feedback of the region to the BO.
3) We experimentally validated using three urinary bladder phantoms requiring different probing maneuvers for acquiring high image quality. The results show that our systems consistently and autonomously acquire high-quality ultrasound images in all phantoms.

We believe that the use of BO combined with domain expertise to perform autonomous ultrasound scanning will lead to less reliance on expert availability and a wider application in remote and underserved populations.

A. Related Work

Autonomous Robotic Ultrasound Systems: In recent years, a range of autonomous robotic ultrasound systems has been proposed to minimize human intervention. Earlier works used image features for ultrasound image-based visual servoing [9]–[11]. Later, various systems used pixel-based confidence map methods [12] and segmentation of structures for optimizing the probe poses and forces [13]–[16]. However, these image feature- and pixel-based approaches are modality specific, computationally expensive and do not consider the significance of diagnostic aspects. Hennersperger *et al.* [17] developed the autonomous system using the pre-operative MRI scan, however, MRI is quite expensive to acquire. Ma *et al.* [18] proposed autonomous lung scanning by localizing the target region using RGB-D sensor data. However, the system used only force feedback and did not rely on ultrasound image feedback, thereby limiting its diagnostic accuracy.

Recently, Li *et al.* [19], [20] proposed a deep Reinforcement Learning (RL) framework to control the probe for spinal ultrasound, incorporating image quality optimization into the reward formulation. However, the success of these systems is limited to phantoms and patients whose data was included during training. Moreover, deploying RL in medical systems is quite challenging, as it requires vast amount of physical interaction with the human body and poses safety and ethical concerns. In contrast to these systems, the proposed autonomous ultrasound system narrows down the area to be scanned using BO, eliminating the need to thoroughly scan the entire region. We further propose using domain expertise gleaned from the experts in the form of BO prior and image quality metrics, in order to acquire diagnostic-quality ultrasound images.

Bayesian Optimization for Medical Robots: Due to the fast optimization capability, BO has been adopted for safety-critical robotic medical procedures, such as autonomous robotic palpation [21], semi-autonomous surgical robot [22], controller tuning of hip exoskeletons [23] and autonomous robotic ultrasound [24], [25]. Our work is a non-trivial extension to the work by Goel *et al.* [25]. They proposed using BO for autonomous ultrasound utilizing segmentation of the vessel in the ultrasound image as feedback to the BO for scanning the region with high vessel density. They used hybrid position-force control to move the robot in (x, y) plane while maintaining constant force along the $z-$direction to the point of contact. In contrast, our work suggests two technical improvements to enhance the practicality of this approach. First, we recommend using a deep learning model that generates quality scores for ultrasound images as feedback to the BO instead of relying on a segmented mask of the tissue or structure. The latter approach can be very time-consuming and labor-intensive for experts as they would need to annotate anatomical structures' boundaries, taking into account the ultrasound image noise and variability due to machine settings, probe pressure, and patient anatomy. Second, we expand the capabilities of the BO by enabling it to search for the optimal scanning region along the (x, y, z)-axis. Notably, the z-axis is under variable force control to account for varying physiological conditions [26].

Domain Expertise in BO: BO can utilize the expert's knowledge in the form of priors (beliefs) that the expert (practitioner) has on the potential location of the optimum. Such techniques have been mostly used for hyper-parameter tuning of image and text datasets [27], open-source machine learning datasets [28] and robot simulation experiments [29]. A few recent works have utilized expert's knowledge in the form of prior for medical robots [30], [31]. Ayvali *et al.* [30] propose robotic palpation to detect tissue abnormalities using BO. They modified the acquisition function of BO, whose value peaks at the user-provided locations. Zhu *et al.* [31] proposed an autonomous robotic auscultation system for locating the optimal sound quality location using BO. They used visual registration of the patient to locate the anatomical landmarks for obtaining a prior observation model. Inspired by these works, we propose BO for autonomous ultrasound leveraging a prior quality map gleaned from expert's demonstrations.

II. Methodology

The pipeline of the autonomous robotic ultrasound system is shown in Fig. 2. In the *offline phase*, the expert will demonstrate the potential probing poses to acquire the diagnostic quality images. This demonstrated data would be used to build a *prior quality map*, which encodes prior anatomical approximation about expected image quality. We also built a dataset of urinary bladder ultrasound images of humans and phantoms with labelled image qualities and trained a deep learning model for image quality assessment metrics. In the *online phase*, we used BO to select the probe poses to find the optimal ultrasound image quality utilizing both the prior map and quality metric gleaned from the domain expertise.

A. Bayesian optimization formulation

We use BO to search adaptively for probing poses that yield a high-quality ultrasound image within a specified

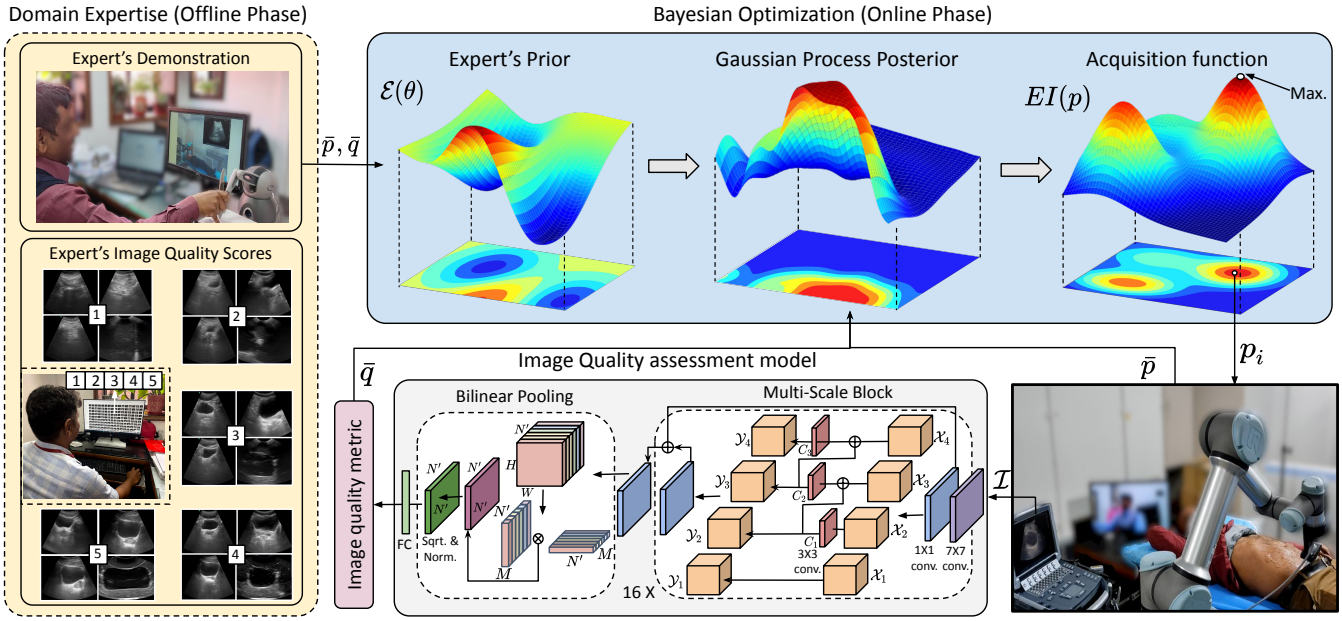

Fig. 2: Overview of the pipeline for autonomous robotic ultrasound using online Bayesian optimization (BO), and offline domain expertise to obtain a prior quality map and to learn image quality assessment metric for providing feedback to BO.

anatomical region. Let A be the region of interest on the human body enclosing the anatomical structure, then the objective of BO is to solve:

$$\max_{\boldsymbol{p}\in A} q(\boldsymbol{\mathcal{I}}(\boldsymbol{p})) \tag{1}$$

where $q(\boldsymbol{\mathcal{I}}(\boldsymbol{p}))$ denotes the quality score of ultrasound Image $\boldsymbol{\mathcal{I}}$ at probe pose $\boldsymbol{p}$. The BO will compute the probabilistic estimate of the unknown quality map $q(\boldsymbol{\mathcal{I}}(\boldsymbol{p}))$ across the human body using the domain expertise in the form of *prior* and *image quality metric*. An *acquisition function* is optimized to yield the new probing pose. Once the new observation is found, the estimate is re-fitted to the data and the process is repeated till the termination criteria is reached, which is either the maximum reasonable iteration N_{max} or the estimated quality score threshold required for adequate diagnosis. The overall algorithm is outlined in Algorithm 1.

1) Expert's prior: A common estimator used in BO is Gaussian Process (GP) model, which defines an unknown function f by assigning a probe pose $\boldsymbol{p}$ a random variable $f(\boldsymbol{p})$, which jointly represent a Gaussian. A GP for unknown function f is defined by the mean function $\boldsymbol{\mu}(\cdot)$ and covariance or kernel function $\boldsymbol{\kappa}(\cdot,\cdot)$. Given the function value estimates $\bar{\boldsymbol{f}} = [f(\boldsymbol{p}_1), \cdots, f(\boldsymbol{p}_n)]$ at probe poses $\bar{\boldsymbol{p}} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n]$, GP regression can predict the function f at new probe pose $\boldsymbol{p}^*$ as the Gaussian distribution and is given by:

$$\mathcal{P}(f(\boldsymbol{p}^*)|\boldsymbol{p}^*, \bar{\boldsymbol{p}}, \bar{\boldsymbol{f}}) = \mathcal{N}(\boldsymbol{k}\boldsymbol{K}^{-1}\bar{\boldsymbol{f}}, \boldsymbol{\kappa}(\boldsymbol{p}^*, \boldsymbol{p}^*) - \boldsymbol{k}\boldsymbol{K}^{-1}\boldsymbol{k}^T) \tag{2}$$

where,

$$\boldsymbol{k} = \begin{bmatrix} \boldsymbol{\kappa}(\boldsymbol{p}_*, \boldsymbol{p}_1) & \cdots & \boldsymbol{\kappa}(\boldsymbol{p}_*, \boldsymbol{p}_n) \end{bmatrix}$$

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

We opted to use a combination of two kernel functions, namely the radial basis function and white noise function, as their combination improved estimations for structures present in ultrasound images [25]. The formulation of the kernel is:

$$\kappa(\boldsymbol{p}_i, \boldsymbol{p}_j) = \sigma_r \exp\left(\frac{-||\boldsymbol{p}_i - \boldsymbol{p}_j||^2}{2l^2}\right) + \sigma_w \boldsymbol{I} \tag{3}$$

where σ_r is the overall variance, l is the length-scale, σ_w is the variance of noise and $\boldsymbol{I}$ is the identity matrix. We further denote the set of image qualities as $\bar{\boldsymbol{q}} = [q_1, \cdots, q_n]$.

In GP, we propose using prior knowledge gleaned from expert's demonstrations to reduce the explorations and capture the variations of probe poses on the magnitude of ultrasound image quality corresponding to different human anatomy. Inspired from work in [31], we formulated the GP as a semi-parametric GP model, with its prior $\mathcal{E}(\boldsymbol{\theta})$ modeled as a Gaussian process with latent parameters $\boldsymbol{\theta}$, representing the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance function $\boldsymbol{\kappa}$. The parameters $\boldsymbol{\theta}$ is initially inferred from observed probe poses and ultrasound image qualities, which the expert will provide by maneuvering the probe at the potential poses of the optimum image quality across different subjects. During online BO, $\boldsymbol{\theta}$ will be inferred using the history of points in $(\bar{\boldsymbol{p}}, \bar{\boldsymbol{q}})$ and prior $\mathcal{E}(\boldsymbol{\theta})$ with Maximum A Posteriori (MAP) estimation, using an L-BFGS solver as:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\bar{\boldsymbol{p}}, \bar{\boldsymbol{q}})\mathcal{E}(\boldsymbol{\theta}) \tag{4}$$

where $\mathcal{L}(\boldsymbol{\theta}|\bar{\boldsymbol{p}}, \bar{\boldsymbol{q}}) = \prod \mathbb{P}(\boldsymbol{q}_i|\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{p}_i), \boldsymbol{K})$ is the likelihood function and $\mathbb{P}(.)$ denotes the probability density function of Gaussian distribution $\mathcal{N}(\boldsymbol{q}_i|\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{p}_i), \boldsymbol{K})$. Since GP models the residual function $f(\boldsymbol{p})$ with respect to the prior, we subtract the prior from image quality as $f(\boldsymbol{p}_i) = q_i - \mu_{\boldsymbol{\theta}}(\boldsymbol{p})$, before re-estimating the GP.

2) Acquisition Function: In each iteration of BO, the next probe pose to observe the image quality is determined using

Algorithm 1: Bayesian Optimization for Ultrasound

Input: Prior $\mathcal{E}(\boldsymbol{\theta})$, Region A, max. iterations N_{max};
Initialize $\bar{\mathbf{p}} = \{\}$, $\bar{\mathbf{f}} = \{\}$, $\bar{\mathbf{q}} = \{\}$;
for $i = 1, ..., N_{max}$ **do**
  $\boldsymbol{p}_i \leftarrow \arg\max_{\boldsymbol{p} \in A} EI(\boldsymbol{p})$;
  if *termination criteria reached* **then**
    stop;
  else
    Probe at $\boldsymbol{p}_i$, compute image quality q_i;
    Set $\bar{\mathbf{p}} \leftarrow \bar{\mathbf{p}} \cup \{p_i\}$, $\bar{\mathbf{p}} \leftarrow \bar{\mathbf{p}} \cup \{q_i\}$;
    $\boldsymbol{\theta} \leftarrow \text{argmax}\, \mathcal{L}(\boldsymbol{\theta}|\bar{\mathbf{p}}, \bar{\mathbf{q}})\mathcal{E}(\boldsymbol{\theta})$;
    Set $\bar{\mathbf{f}} \leftarrow \bar{\mathbf{f}} \cup \{q_i - \mu_{\boldsymbol{\theta}}(\boldsymbol{p})\}$;
    Re-estimate GP;
return Top probe poses with max. image quality;

an acquisition function. We have used an Expected Improvement (EI), which is the most commonly used acquisition function. If the posterior mean and variance of GP is given by $\boldsymbol{\mu}_{\bar{\boldsymbol{f}}}(\boldsymbol{x}), \boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(x)$, then EI can be formulated as:

$$EI(\boldsymbol{p}) = \begin{cases} (\boldsymbol{\mu}_{\bar{\boldsymbol{f}}}(\boldsymbol{p}) - f^+(\boldsymbol{p}) - \xi)\boldsymbol{\Phi}(\boldsymbol{Z}) + \boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(\boldsymbol{p})\boldsymbol{\phi}(\boldsymbol{Z}) & \text{if } \boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(\boldsymbol{p}) > 0 \\ 0 & \text{if } \boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(\boldsymbol{p}) = 0 \end{cases} \quad (5)$$

where $\boldsymbol{Z} = \frac{\boldsymbol{\mu}_{\bar{\boldsymbol{f}}}(\boldsymbol{p}) - f^+(\boldsymbol{p})}{\boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(p)}$ if $\boldsymbol{\sigma}^2_{\bar{\boldsymbol{f}}}(\boldsymbol{p}) > 0$ else 0; $\boldsymbol{\Phi}$ and $\boldsymbol{\phi}$ are the probability and cumulative density function of standard normal distribution, respectively and $f^+(\boldsymbol{p})$ is the best observed quality so far. The parameter ξ in eq. (5) governs the amount of exploration during optimization and a high ξ value leads to more exploration or less exploitation.

B. Expert's ultrasound image quality metric

1) Dataset: We used two datasets of Urinary Bladder (UB) ultrasound images. One of them is collected during the *in-vivo* trials of our in-house developed Telerobotic Ultrasound System [3], [32] at All India Institute of Medical Sciences (AIIMS), Delhi, India. The AIIMS ethics committee approved this study under IEC-855/04.09.2020,RP-16/2020. The other dataset is collected from the UB phantom. A total of 2016 real and 2016 phantom images were collected. The ground truth quality of the images is an average integer score of labels by three expert radiologists, each having 15 years of experience in abdomen radiology. Each label is an integer score between $1-5$, based on an internationally prescribed generalized 5-level absolute assessment scale [5], [33]. A score of 1 means no appearance of the urinary bladder and 5 means that the clear depiction of the urinary bladder with distinct boundaries and acceptable artifacts, depicting a high diagnostic accuracy. A subpar-quality image (2 to 4) either contains noise or motion artifacts, blurred images, indistinct boundaries, obscuring the posterior or anterior sections of the urinary bladder. Later, we normalized the quality score in the range of $0-1$ for standard comparison with other quality estimation methods.

2) Feature extraction: Ultrasound image quality assessment requires rich feature extraction for classifying the images that are highly variable in appearance but differ a lot in terms of image quality, as shown in Fig. 2. In recent work, Song *et al.* [34] proposed a bilinear Convolutional Neural Network (CNN) for fine-grained classification of breast ultrasound image quality. We propose a technical enhancement to this work in order to analyze the urinary bladder ultrasound images, in which the bladder appears at multiple scales/shapes (refer Fig. 2 for sample images) due to the variability among inter- and intra-subject anatomy, probe poses and forces. Thus, it is also essential to analyze images at multiple scales. Recently, Basu *et al.* [35] proposed combining multi-scale and second-order capability for detecting gall bladder cancer. Taking inspiration from these works, we proposed a deep CNN-based quality assessment model. The base network used is Residual Network (ResNet50) [36] and combined with Multi-scale, Bilinear Pooling classifier, as shown in Fig. 2.

We used the group convolution kernels on equal-width feature volume splits in place of the 3×3 convolution kernel in the bottleneck layer of ResNet50. If $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{H \times W \times N}$ represents the feature volume with height H, width W and number of channels N, then the operation of Multi-scale block can be represented by the following equations:

$$\boldsymbol{\mathcal{Y}}_1 = \boldsymbol{\mathcal{X}}_1 \qquad \boldsymbol{\mathcal{Y}}_3 = \boldsymbol{C}_2(\boldsymbol{\mathcal{Y}}_2 + \boldsymbol{\mathcal{X}}_3)$$
$$\boldsymbol{\mathcal{Y}}_2 = \boldsymbol{C}_1(\boldsymbol{\mathcal{X}}_2) \qquad \boldsymbol{\mathcal{Y}}_4 = \boldsymbol{C}_3(\boldsymbol{\mathcal{Y}}_3 + \boldsymbol{\mathcal{X}}_4)$$

where $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{H \times W \times N/4}$. Each split $\boldsymbol{\mathcal{X}}_i$ is first concatenated with output of previous split $\boldsymbol{\mathcal{Y}}_{i-1}$ and then fed to the 3×3 convolutional kernel $\boldsymbol{C}_i$ to produce an output $\boldsymbol{\mathcal{Y}}_i$. After passing the image through 16 multi-scale blocks, the image feature volume $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{H \times W \times N}$ is passed through 1×1 convolution block to reduce the feature volume to $\boldsymbol{X} \in \mathbb{R}^{H \times W \times N'}$. Then it is reshaped to a matrix $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M \times N'}$ where $M = H \times W$. Later, a bilinear pooling is applied as:

$$\boldsymbol{\mathcal{B}} = \frac{1}{N'}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^{\boldsymbol{T}}) + \epsilon \boldsymbol{I} \quad (6)$$

$$\boldsymbol{\mathcal{B}} \leftarrow \frac{\boldsymbol{\mathcal{B}}}{||\boldsymbol{\mathcal{B}}||_2} \leftarrow sign(\boldsymbol{\mathcal{B}})\sqrt{|\boldsymbol{\mathcal{B}}|} \quad (7)$$

where eq. (6) computes the outer product of feature volume, eq. (7) will first perform the element-wise square-root and then the l_2 normalisation of the matrix $\boldsymbol{\mathcal{B}}$. Finally, the flattening of the feature map is done and then a fully connected layer is utilized to return the ultrasound image quality score.

C. Robot control

The robot controller will move the probe to the new pose $\boldsymbol{p} = [x, y, z]$ given by BO, where (x, y) is under position control and z is under force (f_z) control. For the safety of phantoms, the force limits have been set to $20N$ [3]. The orientation of the probe is kept normal to the point of contact. The hybrid position-force control is used for controlling the robot. After searching, the robot may execute the top probe poses with maximum image quality.

III. Results and Discussions

A. Experimental setup

We conducted the experiments on the laboratory setup of the Robotic Ultrasound System at Purdue University, USA,

consisting of a 7-DoF Sawyer collaborative robotic arm (Rethink Robotics, Germany) with Micro Convex MC10-5R10S-3 transducer attached to its end-effector. The US image is captured by the Telemed Ultrasound machine (Telemed Medical Systems, Italy) and is transferred to the laptop. The ultrasound was performed on a urinary bladder phantom (YourDesignMedical, USA). We customized this phantom with the 0.39 inches thick (subjected to manual cutting error) rectangular layers of ballistic gel in order to approximately represent the patient's body with physiological differences. Thus, we present our results using three phantoms, termed as $P0$, $P1$ and $P2$, having 0, 1 and 2 layers, as shown in Fig. 3. The BO and image quality model have been implemented in

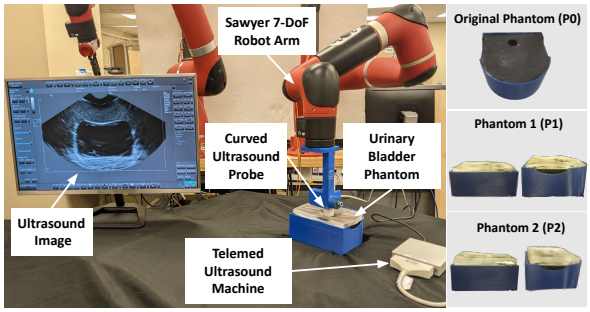

Fig. 3: Experimental setup of robotic ultrasound system with three phantoms of the urinary bladder.

Python 3.8 and PyTorch 1.11. ROS has been used to integrate and establish communication among all components of the setup. For BO Algorithm 1, we used $\xi = 0.1$, $N_{max} = 50$, $A \in ((0, 0.15)m, (0, 0.15)m, (8 - 20)N)$ for (x, y, f_z). The prior $\mathcal{E}(\boldsymbol{\theta})$ has been modeled using GP by fitting it to 10 potential probing poses and corresponding image qualities.

B. Performance of quality assessment model

We trained the ultrasound image quality assessment model, explained in Section II-B, using the Categorical Cross Entropy (CCE) as a loss function. We split the dataset in a $90:10$ ratio as a training and testing dataset. We also used the transfer learning approach [37] and used the proposed model pre-trained on ImageNet. The stochastic gradient descent has been used as an optimizer with a learning rate of 0.005, momentum of 0.9 and weight decay of 0.0005. The size of the input image to the model is 224×224, batch size is 16 and the network is trained for 100 epochs. The results in Table I shows that the proposed model (ResNet50+MS+BP) achieved an increase in accuracy by 3.01% on a test set when compared to the ResNet50+BP model proposed in [34].

C. Comparing different BO strategies

In order to analyze the effectiveness of the proposed methodology, we have compared the BO with zero prior to the BO with the proposed expert's prior. We illustrated these search strategies using the image feedback having a mean of the segmented mask of the bladder in the ultrasound image (q_S) as used in [25] and having proposed ultrasound image quality metric learned from expert's rating (q_E). For segmentation, we used a U-net-based segmentation model

TABLE I: Comparison of ultrasound image quality assessment model predictions on testing set, where accuracy values close to 100 indicate similarity to the expert's quality score.

Image quality score	ResNet50+BP [34]			ResNet50+MS+BP (Proposed)		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
1	92.00	97.87	94.84	93.88	97.87	95.88
2	83.33	68.96	75.47	83.87	89.65	86.67
3	65.22	75.00	69.77	88.24	75.00	81.08
4	93.33	85.71	89.36	91.11	83.67	87.23
5	90.48	97.44	93.83	90.48	97.44	93.83
Average	**87.67**	**87.48**	**87.34**	**90.23**	**90.17**	**90.05**

proposed in [38]. Further, each of the feedback strategies has been compared with different search spaces, first considering the probe motion along x and y-axis of the phantom, second along x, y and $z-$axis of the phantom, where $z-$axis is under the force control (f_z). The estimated quality maps obtained using these strategies for $P0$ are shown in Fig. 4, where red region shows the high-quality region and blue region shows the low-quality region. The black dots over the map represents the queried probe positions over the phantom during the optimization. The first column in Fig. 4 shows the quality map obtained using the uniform movement of the probe over the phantom, which has been considered as the approximate ground truth quality map. For both the quality types, the ground truth has been obtained using the approximate desired force (f_d) of $14N$, $16N$ and $18N$ for $P0$, $P1$ and $P2$, respectively, which gives the best image quality in these phantoms. We present results for 3 cases to illustrate the effect of searching with appropriate force in these phantoms: (i) $f_z < f_d$: fz is constant but equal to $f_d - 4$, (ii) $f_z = f_d$ and (iii) when f_z is variable.

We compared the quality maps of these strategies by doing quantitative analysis using three metrics: (i) Sum of quality difference of top n points, (ii) Top quality, and (iii) Zero Normalized Cross Correlation (ZNCC), as shown in Table II. The numbers in the table represent the average value of the matrices for the 3 tests on each phantom. These metrics have been computed with respect to the approximate ground truth for the phantom. The sum of the difference between the top $n-$points compares the quality of images acquired from the top-n highest quality values, top quality compares the highest value of image quality score and ZNCC evaluates the overall similarity of the acquired quality map during the search. The value of quality differences close to 0, and top quality and ZNCC value close to 1 indicates a better estimation of the quality map. The quality maps in Fig. 4 with less scattered probe points (less exploration) and more points in the high-quality region (red) represent a better search strategy.

From the result in Fig. 4 and Table II, it has been found that the BO using the segmented image as quality score in (x, y) space with $f_z \leq f_d$ have resulted in being too exploratory (low ZNCC) with a lot of points spread over the low-quality region of the phantom. However, the quality maps obtained using the expert's quality metric of the image have fewer explorations, with most of the probe positions in the high-quality region of the phantom. Due to noise and shadows in the ultrasound image, the segmentation results

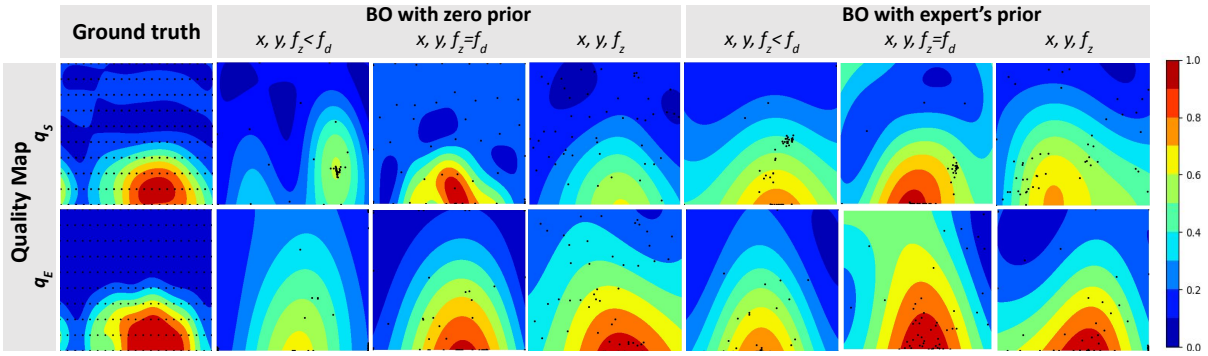

Fig. 4: The estimated ultrasound image quality map of urinary bladder phantom $P0$ using different BO strategies. Black dots are positions where probe evaluated the quality. The corresponding ultrasound images are available in the attached media.

TABLE II: Quantitative comparison of different BO strategies for three different urinary bladder phantoms $P0, P1$ and $P2$

Image quality estimation method	Variables	BO with zero prior						BO with expert's prior					
		Sum of quality difference of n points				Top Quality	ZNCC	Sum of quality difference of n points				Top Quality	ZNCC
		$n=1$	$n=5$	$n=10$	$n=20$		$n=50$	$n=1$	$n=5$	$n=10$	$n=20$		$n=50$
Segmentation	$x, y, f_z < f_d$	0.382	0.944	1.430	1.931	0.684	0.689	0.291	0.692	0.941	1.205	0.782	0.782
(q_S)	$x, y, f_z = f_d$	0.132	0.609	0.963	1.651	0.911	0.811	0.103	0.531	0.785	1.308	0.911	0.920
[32]	x, y, f_z	0.396	1.016	1.919	3.336	**0.711**	**0.733**	0.404	0.991	1.379	2.158	**0.799**	**0.801**
Expert's image	$x, y, f_z < f_d$	0.280	0.370	0.600	1.030	0.690	0.717	0.130	0.290	0.570	1.010	0.750	0.817
quality metric	$x, y, f_z = f_d$	0.120	0.240	0.390	0.710	0.950	0.876	0.050	0.090	0.180	0.970	0.980	0.959
(q_E)	x, y, f_z	0.130	0.270	0.820	1.320	**0.823**	**0.821**	0.040	0.280	0.760	1.600	**0.910**	**0.889**

are prone to errors, resulting in a large number of probe evaluations in low-quality regions, whereas expert's image quality score, which is based on the holistic assessment of the image, pinpoints the focus on anatomical structures rather than getting distracted by noise. The search strategies using $fz < f_d$ could not find the high-quality region and instead converged to the local maxima rather than the global maxima. However, with $f_z = f_d$, the high-quality regions have been acquired. When the quality region is searched using f_z as a variable in BO with zero prior, the quality maps and top quality score show that the high-quality regions can be located with a varying force too, which is essential for in-human ultrasound procedures. However, the search is quite exploratory, reporting low ZNCC values of 0.733 and 0.821 for quality q_S and q_E, respectively. When the expert's prior is used, all BO strategies have significantly improved, including the search space with three variables (x, y, f_z). The exploration steps of BO usually increase as the search space dimension increases. However, BO with expert's prior reported a top quality of 0.910 with a ZNCC score of 0.889, which is 9.6% and 7.6% more than the BO with zero prior.

D. Validating the convergence of probe positions and forces

Since our study involves phantom experiments, the approximate probe positions and forces that yield the best-quality images are known. The search strategy should converge to these approximate probe poses and forces to acquire high-quality images. The proposed strategy has reached the desired probe position with an average mean value accuracy of 98.73% across all phantoms. To emphasize the convergence of force, we compared the probe forces explored by different BO search strategies, as shown in Fig. 5. The proposed formulation of BO using the expert's prior and image quality metric has resulted in the mean value accuracy

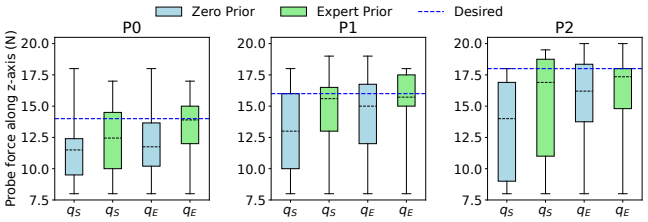

Fig. 5: Force f_z profiles with different BO search strategies

of 99.28%, 98.25%, and 96.11% for $P0$, $P1$, and $P2$, respectively. Comparatively, the other BO search strategies using zero-prior and segmentation-based quality maps (q_S) have shown significant errors in mean values and greater standard deviation due to the noise in image feedback and the inability to adapt to the profile of the scanning region.

IV. Conclusion

We proposed an autonomous Robotic Ultrasound System (RUS) to perform the ultrasound as per clinical protocols. We used Bayesian Optimization (BO) to search for high-quality regions leveraging the domain expertise in the form of a prior quality map and ultrasound image quality. The prior map has been gleaned using expert's demonstration of the potential high-quality probing maneuvers. A novel image quality metric has been learned from the expert-labelled dataset of ultrasound images. Three phantom experiments validated that incorporating domain expertise into BO effectively improves the system performance, resulting in acquiring diagnostic quality ultrasound images while adapting to desired probing maneuvers. Since phantom results are promising, we would like to validate its capability for *in-vivo* study using our RUS in India [3], which is our future work. We would also expand the search space in BO from $[x, y, f_z]$ to include $[roll, pitch, yaw]$ in order to orient the probe for scanning patients with complex physiological conditions.

REFERENCES

[1] A. Rykkje, J. F. Carlsen, and M. B. Nielsen, "Hand-held ultrasound devices compared with high-end ultrasound systems: a systematic review," *Diagnostics*, vol. 9, no. 2, p. 61, 2019.

[2] S. J. Adams, B. Burbridge, H. Obaid, G. Stoneham, P. Babyn, and I. Mendez, "Telerobotic sonography for remote diagnostic imaging: narrative review of current developments and clinical applications," *Journal of Ultrasound in Medicine*, vol. 40, no. 7, pp. 1287–1306, 2021.

[3] 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.

[4] J. Wang, C. Peng, Y. Zhao, R. Ye, J. Hong, H. Huang, and L. Chen, "Application of a robotic tele-echography system for covid-19 pneumonia," *Journal of Ultrasound in Medicine*, vol. 40, no. 2, pp. 385–390, 2021.

[5] 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.

[6] S. J. Adams, B. E. Burbridge, A. Badea, L. Langford, V. Vergara, R. Bryce, L. Bustamante, I. M. Mendez, and P. S. Babyn, "Initial experience using a telerobotic ultrasound system for adult abdominal sonography," *Canadian Association of Radiologists' Journal*, vol. 68, no. 3, pp. 308–314, 2017.

[7] E. Rojas-Muñoz and J. P. Wachs, "Assessing task understanding in remote ultrasound diagnosis via gesture analysis," *Pattern Analysis and Applications*, vol. 24, no. 4, pp. 1489–1500, 2021.

[8] K. Li, Y. Xu, and M. Q.-H. Meng, "An overview of systems and techniques for autonomous robotic ultrasound acquisitions," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 2, pp. 510–524, 2021.

[9] R. Mebarki, A. Krupa, and F. Chaumette, "2-d ultrasound probe complete guidance by visual servoing using image moments," *IEEE Transactions on Robotics*, vol. 26, no. 2, pp. 296–306, 2010.

[10] R. Nakadate, J. Solis, A. Takanishi, E. Minagawa, M. Sugawara, and K. Niki, "Out-of-plane visual servoing method for tracking the carotid artery with a robot-assisted ultrasound diagnostic system," in *2011 IEEE International Conference on Robotics and Automation*, pp. 5267–5272, IEEE, 2011.

[11] C. Nadeau, A. Krupa, J. Petr, and C. Barillot, "Moments-based ultrasound visual servoing: From a mono-to multiplane approach," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1558–1564, 2016.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, and N. Navab, "Towards mri-based autonomous robotic us acquisitions: a first feasibility study," *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 538–548, 2016.

[18] X. Ma, Z. Zhang, and H. K. Zhang, "Autonomous scanning target localization for robotic lung ultrasound imaging," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9467–9474, IEEE, 2021.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] Z. Wang, G. E. Dahl, K. Swersky, C. Lee, Z. Mariet, Z. Nado, J. Gilmer, J. Snoek, and Z. Ghahramani, "Pre-training helps bayesian optimization too," *arXiv preprint arXiv:2207.03084*, 2022.

[28] C. Hvarfner, D. Stoll, A. Souza, M. Lindauer, F. Hutter, and L. Nardi, "pibo: Augmenting acquisition functions with user beliefs for bayesian optimization," *arXiv preprint arXiv:2204.11051*, 2022.

[29] Z. Wang, B. Kim, and L. P. Kaelbling, "Regret bounds for meta bayesian optimization with an unknown gaussian process prior," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[30] E. Ayvali, A. Ansari, L. Wang, N. Simaan, and H. Choset, "Utility-guided palpation for locating tissue abnormalities," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 864–871, 2017.

[31] 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.

[32] 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.

[33] P. Cantin and K. Knapp, "Diagnostic image quality in gynaecological ultrasound: Who should measure it, what should we measure and how?," *Ultrasound*, vol. 22, no. 1, pp. 44–51, 2014.

[34] 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.

[35] S. Basu, M. Gupta, P. Rana, P. Gupta, and C. Arora, "Surpassing the human accuracy: Detecting gallbladder cancer from usg images with curriculum learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20886–20896, 2022.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[37] P. M. Cheng and H. S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," *Journal of digital imaging*, vol. 30, no. 2, pp. 234–243, 2017.

[38] 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.