

Semi-Autonomous Robotic Assistance for Gallbladder Retraction in Surgery

Alexander Schüßler^{2,**}, Christian Kunz^{1,**}, Rayan Younis³, Benjamin Alt⁴, Jamie Paik², Martin Wagner³, Franziska Mathis-Ullrich^{1,*}

Abstract—(Semi-)autonomous robotic assistance in minimally invasive surgery has the potential to alleviate surgical staff shortage and decrease the workload of medical professionals. These robots must execute complex tasks within unpredictable and unstructured environments encountered during surgery. Although imitation learning approaches have the potential to learn complex surgical skills, the interpretation of robot behavior during safety-critical scenarios, such as surgery, remains a challenge. Through combining interpretable 3D point cloud feature vectors based on domain knowledge with feedforward neural networks and probabilistic movement primitives, domain knowledge-informed movement primitives effectively learn surgical skills while improving the interpretation of robot behavior. The evaluation on test data proves that the proposed method can effectively learn surgical skills based on a small number of demonstrations. Using the proposed imitation learning method, a semi-autonomous robotic assistance for directed gallbladder retraction is introduced and evaluated during gallbladder removal on a silicone liver phantom and *ex vivo* porcine livers. Achieving over 91 % and 92 % successful gallbladder retractions, the robotic assistance enables effective support for surgeons during these surgical interventions.

Index Terms—Surgical Robotics, Informed Machine Learning, Imitation Learning, Learning from Demonstration, Minimally Invasive Surgery.

I. INTRODUCTION

ROBOT-assisted surgery with (semi-)autonomous robotic assistance systems has the potential to address surgical staff shortages and surgeon workload on a wide scale by supporting the surgical team [1]–[11]. Robotic assistance for task autonomy, i.e. level of autonomy (LoA) 2, in which surgeons and robots collaborate intraoperatively, propose an alternative to robotic surgery with full autonomy (LoA 5), which remains challenging due to task complexity, environmental difficulties, and the degree of human involvement in surgical procedures.

* Corresponding author: franziska.mathis-ullrich@fau.de.

** Authors contributed equally.

¹ Surgical Planning and Robotic Cognition Laboratory, Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-University Erlangen-Nuremberg, Germany.

² Reconfigurable Robotics Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

³ Department of Visceral, Thoracic and Vascular Surgery, Faculty of Medicine, University Hospital Carl Gustav Carus & Center for the Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, Germany.

⁴ ArtiMinds Robotics GmbH, Karlsruhe, Germany.

This work was supported by the German Federal Ministry of Education and Research under the grant 13GW0471 and by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence “Centre for Tactile Internet with Human-in-the-Loop” (CeTI) of Technische Universität Dresden.

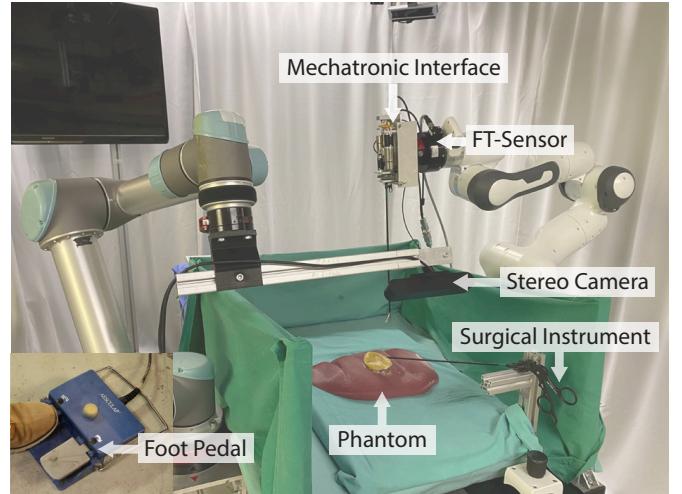


Fig. 1. Experimental setup for gallbladder retraction during cholecystectomy with the semi-autonomous robotic assistance. A collaborative robot with a mechatronic interface for controlling surgical graspers is used to perform gallbladder retractions.

Gallbladder removal, i.e. cholecystectomy, is a common surgical procedure to remove the gallbladder, which requires diverse surgical tasks, such as dissection, clipping, cutting, and retraction, to separate the gallbladder from the liver bed. The procedure is commonly performed by a surgeon who is supported by a surgical assistant [12]. Thus the procedure is ideal for introducing (semi-)autonomous robotic assistance systems. Previous works have automated surgical tissue retraction without translation to *ex vivo* tissue [11], [13], [14]. Liang et al. [15] introduce autonomous retraction for tissue dissection on *ex vivo* chicken tissue, but did not evaluate the method for cholecystectomy. Moreover, Oh et al. [16] focus on autonomous recognition of the dissection boundary on an *ex vivo* porcine gallbladder. However, the work uses pre-defined path planning of the retraction direction.

To execute diverse surgical tasks, the robots need to interact with the complex and dynamic environment of the surgical scene consisting of deformable organs and understand the surgeon’s actions and intentions to provide useful assistance. Learning-based methods such as imitation learning show a promising approach to achieve robot behavior that addresses these challenges of robotic surgery. Imitation learning allows robots to learn surgical skills from a small number of human demonstrations. To date, commonly used imitation learning approaches for surgical procedures include generative

adversarial imitation learning [17], adaptive model predictive control [13], and movement primitive-based approaches [18]–[20]. Movement primitive-based approaches, such as Dynamic movement primitives (DMP) [18], deep movement primitives (DeepMP) [21], and probabilistic dynamic movement primitives (ProDMP) [20], combine basic robot movements based on human demonstrations into more complex robot skills. However, these methods often rely on an end-to-end learning process, in which the input and output features of the robot motion generation are difficult to interpret, which is especially important in the context of certification and acceptance of robotic assistance in surgery.

In order to make the input and output of autonomous assistance systems understandable for surgeons, previous works have introduced domain knowledge into the systems' decision process to improve their interpretability [22], [23]. Meli et al. [22] introduce interpretable task planning and situation awareness of an autonomous robotic assistance system by encoding prior expert knowledge in a logic module and providing semantic interpretation of sensor data with a patient-specific biomechanical simulation. Including domain knowledge in imitation learning approaches holds significant promise for achieving (semi-)autonomous robotic assistance with complex and interpretable behavior. However, robotic assistance for surgeons that combines efficient and accurate robot movement prediction with interpretability of the prediction remains a challenge.

In this work, we introduce a semi-autonomous robotic assistance system with task autonomy (LoA 2) for the directed gallbladder retraction task in cholecystectomy on phantom and *ex vivo* scenario. We introduce domain knowledge-informed movement primitives (DKMP) that combine interpretable 3D point cloud feature vectors based on domain knowledge with feedforward neural networks and probabilistic movement primitives (ProMP) trained on expert demonstrations. Our method is used to learn a movement primitive for performing gallbladder retraction in the desired direction depending on the current surgical scene. We validate and compare DKMP on datasets with unknown gallbladder retractions on a surgical liver phantom [24] and *ex vivo* porcine livers. The results indicate that DKMP can accurately predict gallbladder retractions, while having interpretable input and output features. Finally, we deploy DKMP as a semi-autonomous robotic assistance system for directed gallbladder retraction tasks in phantom and *ex vivo* surgeries, resulting in a success rate of 91 % and 92 %, respectively.

The main contributions of this work can be summarized as:

- A semi-autonomous robotic assistance system for directed gallbladder retraction tasks in phantom and *ex vivo* cholecystectomies.
- Domain knowledge-informed movement primitives (DKMP), an interpretable imitation learning method for learning surgical tasks from expert demonstrations by making explicit use of domain knowledge.
- An experimental evaluation of DKMP on phantom and *ex vivo* porcine liver surgeries.

II. METHODS

We propose domain knowledge-informed movement primitives (DKMP) to incorporate domain knowledge in the input features of robot movement prediction. DKMP combines a domain knowledge-based feature extraction from point clouds, a neural network-based movement prediction and a ProMP-based trajectory generation mechanism to predict gallbladder retraction trajectories based on point clouds of the current surgical scene (Figure 2).

A. Feature Extraction

To extract interpretable features from the point cloud of the current surgical scene, DKMP includes a feature extraction step (Figure 2). The definition of these extracted features is crucial for the efficacy and interpretability of DKMP, as they serve as build the input features for the movement prediction neural network. Therefore, the definition is based on domain knowledge of the surgeons from the clinical practice. While this work presents the feature extraction for the gallbladder retraction during cholecystectomy, a similar feature extraction process could be used for other surgical tasks.

To ensure the safe dissection of the gallbladder during cholecystectomy, it is essential to apply the appropriate tension at the specific location where the next dissection step will be performed. This location, referred to as the *next dissection point* p_{next} , lies on the accessible section of the dissection plane, also known as the *dissection line* (Figure 3). The correct gallbladder retraction direction and force depend on the location of the *next dissection point* p_{next} , as well as the progress of the dissection line. The retraction retracting direction of the gallbladder in relation to the endoscope is mostly facing ventrally (upwards, z-axis) and cranially (backwards, y-axis) to the patient's left, right or middle (x-axis) depending on the location of the *next dissection point* p_{next} along the *dissection line* (Figure 3).

Using this knowledge of the clinical practice, we define five distinct point cloud landmarks in the surgical scene of the gallbladder removal:

- The *lowest point* of the liver bed p_{start} ,
- *left-most point* of the *dissection line* p_{left} ,
- *right-most point* of the *dissection line* p_{right} ,
- *next dissection point* shown by the surgeon with the tip of the surgical instrument p_{next} ,
- and *grasping point* defined by the tip of the surgical grasper p_{grasp} .

These landmarks are distinct in the scene of both the surgical liver phantom (Figure 3a), as well as the *ex vivo* porcine liver (Figure 3b). The lowest point of the liver bed p_{start} is used as a reference point for the other landmarks, while the landmarks p_{left} , p_{right} and p_{next} represent the current *dissection line*. The current contact location of the surgical grasper is represented by the landmark p_{grasp} , which is extracted in the center of the grasper tip. At each step of the surgical procedure, the surgeon manually identifies key landmarks in the point cloud (as depicted in Fig. 3c) through an intuitive graphical user interface (task autonomy, LoA 2 [5]). While the focus of this paper is automated trajectory generation, future work will aim

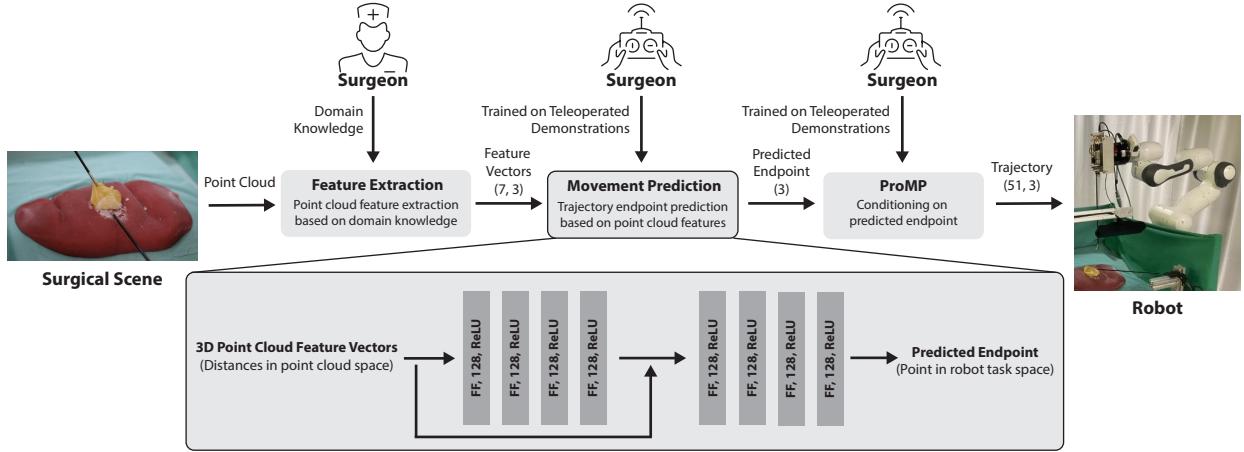


Fig. 2. Domain knowledge-informed movement primitives (DKMP): DKMP combine the extraction of interpretable features from point clouds based on domain knowledge with neural network-based movement prediction, and probabilistic movement primitives (ProMP) to generating gallbladder retraction trajectories depending on the current surgical scene.

plans to automate the extraction of these landmarks from the point cloud, giving the surgeon task strategies to select from (conditional autonomy, LoA 3 [5]).

The defined point cloud landmarks are the basis for the interpretable input features of the DKMP. Based on these five point cloud landmarks, we compute seven 3D feature vectors, each capturing the spatial relationship between the landmarks of our surgical scene. Stacking these vectors yields a $(7,3)$ matrix, which serves as input to the neural network for movement prediction:

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \end{pmatrix} = \begin{pmatrix} p_{\text{left}} - p_{\text{start}} \\ p_{\text{right}} - p_{\text{start}} \\ p_{\text{next}} - p_{\text{start}} \\ p_{\text{grasp}} - p_{\text{start}} \\ p_{\text{left}} - p_{\text{next}} \\ p_{\text{right}} - p_{\text{next}} \\ p_{\text{grasp}} - p_{\text{next}} \end{pmatrix} \in \mathbb{R}^{7 \times 3}. \quad (1)$$

The 3D point cloud feature vectors f_1, f_2, f_3 , and f_4 provide information about the current progress of the gallbladder separation by computing the distance of the *dissection line* p_{left} , the *next dissection point* p_{next} , and the *grasping point* p_{grasp} to the *lowest point* p_{start} of the liver bed (which indicates the starting position of the separation). Additionally, the 3D point cloud feature vectors f_5 and f_6 contain information about the location of the surgical instrument p_{next} on the *dissection line*. Thus, they give information on where the surgeon will perform the next dissection, which is crucial to understand the desired direction of the gallbladder retraction. Assuming consistent material properties of each gallbladder and a linear spring relationship, the 3D point cloud feature vector f_7 delivers information about the current tension of the gallbladder at the dissection line. This is important to understand the current gallbladder tension and the necessary retraction movement to achieve the desired tension at the *next dissection point*.

B. Movement Prediction and Generation

To generate a robot trajectory based on the extracted 3D point cloud feature vectors, DKMP utilizes a neural network-based movement prediction and ProMP to generated the gallbladder retraction trajectory (Figure 2).

The movement prediction uses the extracted 3D point cloud feature vectors, based on domain knowledge, as its input features. We use a feedforward neural network to predict the target point of the gallbladder retraction trajectory. The feedforward neural network g_θ maps the 3D point cloud feature vectors $\mathbf{f} = (f_1, f_2, f_3, f_4, f_5, f_6, f_7)^T$ to the gallbladder retraction endpoint $\mathbf{p}_{\text{end}} = g_\theta(\mathbf{f})$. We use a feedforward neural network consisting of 8 layers with 128 neurons and ReLU activations. The layers are divided into two blocks of 4 layers with a skip connection added between the two blocks. The skip connection is similar to the ones used in the ResNet architecture, which improve the optimization process for deeper neural networks [25]. Using the collected datasets of gallbladder retraction demonstrations on the phantom and *ex vivo* porcine livers, the weights θ of the feedforward neural network g_θ are trained to convergence by backpropagation using the Adam optimizer with an initial learning rate of 0.0001 and a step-wise reduction schedule. We use the mean squared error (MSE) loss over all N trajectories between the predicted gallbladder retraction trajectory endpoint $\mathbf{p}_{\text{end,p}}$ and the desired gallbladder retraction trajectory endpoint $\mathbf{p}_{\text{end,d}}$ for training:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_{\text{end,p}} - \mathbf{p}_{\text{end,d}}\|^2. \quad (2)$$

To generate a complete trajectory for executing the gallbladder retraction with a robot, we utilize ProMP [26], [27]. ProMP represent movement primitives as probabilistic distributions over the demonstrated trajectories, which enables the utilization of probabilistic operations for improved generalization. We use a ProMP with 30 normalized Gaussian basis functions (10 for each dimension) to represent the gallbladder retraction trajectory. Using the gallbladder retraction demonstrations of

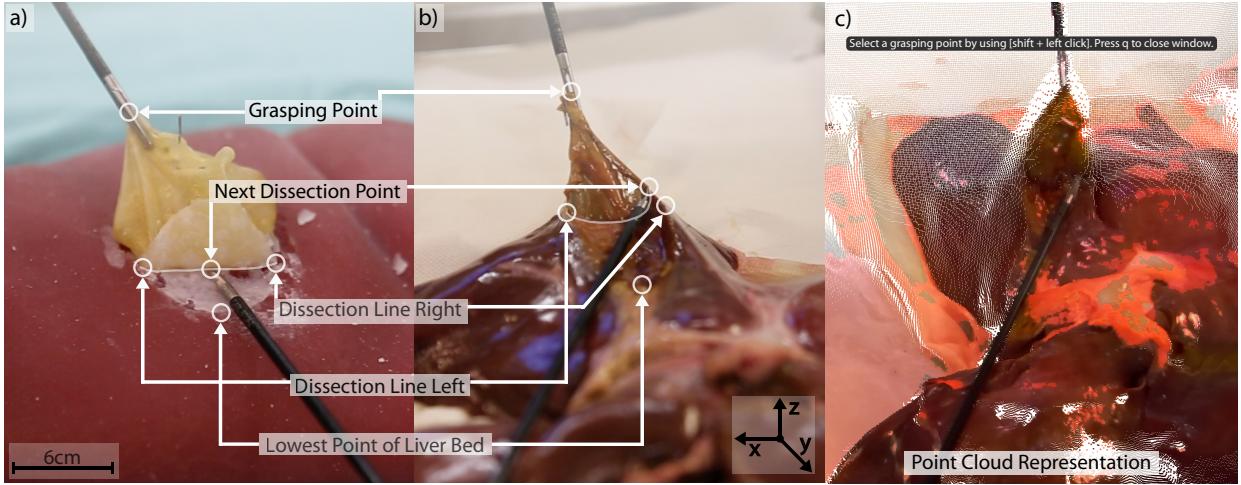


Fig. 3. The point cloud landmarks of the gallbladder retraction are selected based on domain knowledge by the clinical expert. The surgical landmarks include the *lowest point* of the liver bed, the *left-most point* and *right-most point* of the *dissection line*, the *next dissection point* on the *dissection line*, and the *grasping point* on the phantom (a) and *ex vivo* (b) gallbladder. The surgical landmarks are manually extracted from the point cloud using a graphical user interface (c).

the datasets, we train the weights for the 30 basis functions through maximum likelihood estimation using the Expectation Maximization algorithm [28]. In order to generalize the gallbladder retraction to predicted trajectory endpoints $p_{\text{end},p}$ of the movement prediction, we utilize the conditioning ability of ProMP to modulate the gallbladder retraction trajectory.

III. EXPERIMENTAL SETUP

A. Surgical Setup

To realize and validate the semi-autonomous robotic assistance for cholecystectomy, we introduce a surgical setup for phantom and *ex vivo* surgeries. A critical activity during cholecystectomy involves retracting the gallbladder to ensure its safe separation from the liver bed. Poor tension results in bad exposure, that may lead to mistakenly dissecting into the liver or gallbladder, potentially causing liver bleeding or gallbladder perforation. Both could result in relevant complications, such as postoperative bleeding and infection. Traditionally, this retraction task is performed by a surgical assistant with one hand while using the other hand to control the endoscope (i.e., intra-abdominal camera). We develop a semi-autonomous robotic assistance for this retraction task, while the dissection task is still performed by the surgeon. The robotic assistance is activated by the surgeon and uses visual information to predict and execute the gallbladder retraction for tensioning the gallbladder.

For the implementation and evaluation of the semi-autonomous robotic assistance, we simplify the minimally invasive *in vivo* setup. The surgeries are performed on a surgical liver phantom and *ex vivo* porcine livers. The phantom [24] uses silicone-based tissue and latex gallbladders to replicate the gallbladder retraction and separation on *ex vivo* porcine livers. While the phantom enables the same surgical task as *ex vivo* porcine livers, the retraction forces of the latex gallbladders vary in comparison to *ex vivo* gallbladders. In order to translate the imitation learning methods to real

tissue, we use fresh-frozen *ex vivo* porcine livers from the local abattoir with the gallbladder still attached. The gallbladder is separated using Maryland scissors for the phantom and *ex vivo* scenario. The surgical setup used in this work has no motion constraints, in contrast to the remote center of motion setup of minimally invasive surgery. To capture the necessary point cloud for manual landmark determination we use a stereo camera (ZED2, Stereolabs Inc., USA). These simplifications are a result of compromising between a setup for fast data collection and a realistic surgical scenario, as the focus of this work is the validation of the proposed DKMP method and its usage to learn surgical tasks.

B. Robotic Setup

To collect the necessary data for training and evaluating the DKMP method, we introduce a setup based on two collaborative robotic arms, as presented in Figure 1. The setup includes a stereo camera (ZED2, Stereolabs Inc., USA) attached to a stationary robotic arm (UR5, Universal Robots, Denmark) to provide static and consistent visual information about the surgical scene. We use the second robotic arm (Franka Panda version 1, Franka Robotics GmbH, Germany) with an attached 6 degrees of freedom (DoF) force-torque sensor (Koris Force & Safety Components GmbH, Germany) and a custom-made mechatronic interface to control conventional laparoscopic instruments, here a grasper (Karl Storz SE & Co. KG, Germany) to retract the gallbladder. We implemented a teleoperation mode to collect expert demonstrations using an Xbox Controller (Microsoft, USA) and an autonomous mode for trajectory determination and execution with our proposed method. While gallbladder retraction is performed autonomously using a robotic assistance system, gallbladder separation is performed manually by the surgeon. Using this setup, we perform gallbladder removal surgeries on a surgical liver phantom and *ex vivo* porcine livers. Overall, this robotic setup enables extensive data collection for training and evaluation of the imitation learning methods.

C. Dataset Acquisition

To train and evaluate our proposed method for gallbladder retraction, custom datasets with teleoperated demonstrations on the phantom and *ex vivo* porcine livers are created. We utilize the experimental setup (Figure 1) for the data collection process. Each demonstration includes trajectories with position and force information, as well as visual information (RGB image and point cloud) about the surgical scene captured from the stereo camera. To collect the tool position and force trajectories of the retraction demonstrations, we use the internal encoders of the Franka Panda robot arm and the external 6 DoF force-torque sensor. We collect two separate datasets with gallbladder retractions based on one phantom with latex gallbladder and two *ex vivo* porcine livers. The gallbladder retraction demonstrations were performed by technical research staff on the basis of explanations and demonstrations of surgeons. We collected retraction demonstrations in different variations, throughout the gallbladder removal process from fully attached to nearly separated gallbladders.

Positions for the *next dissection point* on three different areas along the *dissection line* were defined: left, middle and right. This simplification was done to cluster the demonstrations. The *next dissection points* were distributed over these areas. During data collection, we ensured uniform distribution over these three classes to reduce the intrinsic bias of the dataset towards a certain *next dissection point* or area.

Based on this data collection procedure, the phantom dataset includes 252 demonstrated gallbladder retractions (from one phantom), while the *ex vivo* dataset contains 270 demonstrations (from two *ex vivo* livers).

Both datasets contain the points of the trajectories, retraction forces, RGB images, and RGB point clouds. The datasets are each divided into a training set D_{train} , validation set D_{val} , and test set D_{test} , following a split ratio of 78.6%, 10.7%, and 10.7% for the phantom dataset and 73.4%, 13.3%, and 13.3% for the *ex vivo* dataset. The data collected from the two *ex vivo* porcine livers is equally distributed in training, test, and validation set. The splits are a combined result of the number of demonstrations and the uniform distribution over the three classes of *next dissection points*.

D. Experimental Procedure on Test Datasets

To validate the proposed DKMP method, we evaluate the gallbladder retraction prediction accuracy (both trajectory points and endpoint) on phantom and *ex vivo* porcine liver test datasets and compare it to the prediction of these points using deep movement primitives (DeepMP) [21].

We compute the mean euclidean distance between the predicted and demonstrated (i.e. desired) trajectory endpoints p_{end} (Equation 3) and trajectories τ (size T) (Equation 4) of all test dataset samples N . In addition, we calculate the standard deviation of the euclidean distance. DeepMP use end-to-end learning for mapping visual information directly to robot trajectories. In contrast to the 3D point cloud feature vectors in DKMP, which rely on domain knowledge, DeepMP uses a convolutional autoencoder to extract a feature space from the RGB image of the surgical scene. We use DeepMP

as the baseline for the evaluation to prove that DKMP's prediction performance is comparable to current end-to-end based movement primitive methods.

The accuracy metrics provide an initial assessment of our method's performance, although precisely defining an endpoint position is challenging. From a medical standpoint, it is more about indicating the direction in which the gallbladder must be stretched. Therefore, in addition to the prediction accuracies, we evaluate DKMP's prediction performance based on the retraction directions. This evaluation is carried out by a medical expert based on the plot results depicted in Fig. 4 a-c. Successful retractions direct cranially, i.e. towards the back of the gallbladder, which results in tension of the gallbladder at the *dissection line*. Further, the retraction directions (left, middle, or right) are distinguished depending on the location of the *next dissection point* on the *dissection line*. A *next dissection point* in the middle of the *dissection line* should lead to a retraction direction towards the middle of the gallbladder, while a *next dissection point* on the left should lead to a retraction direction to the right and vice versa for the right side. In addition to evaluating each method on the same scenario that it was trained on, we also evaluate the performance of the phantom-trained methods on the *ex vivo* dataset.

$$\mathbf{p} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_{\text{end},\mathbf{p},i} - \mathbf{p}_{\text{end},\mathbf{d},i}\|_2 \quad (3)$$

$$\tau = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{j=1}^T \|\tau_{\text{end},\mathbf{p},i,j} - \mathbf{p}_{\text{end},\mathbf{d},i,j}\|_2 \quad (4)$$

E. Experimental Procedure in Phantom and Ex Vivo Surgeries

As a final evaluation, we apply the semi-autonomous assistance with DKMP to phantom and *ex vivo* porcine liver cholecystectomies. We use the same experimental setup as the one used for the dataset collection (Figure 1). Using the setup, a medical expert with 3 years of experience as a first surgical assistant of cholecystectomies performed five surgeries on the phantom and four surgeries on *ex vivo* porcine livers. Additionally, one technical expert with 5 years of experience with the cholecystectomy procedure conducted a single *ex vivo* surgery. During the surgeries, the gallbladder retractions were performed using the semi-autonomous assistance, while the other steps of the cholecystectomy were performed manually by the clinical and technical expert. The assistance was initiated by the surgeon through a foot pedal (Figure 1) and the surgical landmarks for calculation of the 3D point cloud feature vectors were manually selected in the point cloud using a graphical user interface (Fig. 3c). In addition to the trajectory execution, we added a safety-force limit of 4 N for the phantom surgeries and 2.5 N for the *ex vivo* surgeries to the controller of the robotic arm, which we determined in previous works [24], based on the recorded datasets.

Further, the system continues along the trajectory until the force limit is reached, which makes it sometimes necessary to pull beyond the predicted endpoint, following a retraction vector. The retraction vector is constructed based on the vector between the last two trajectory points.

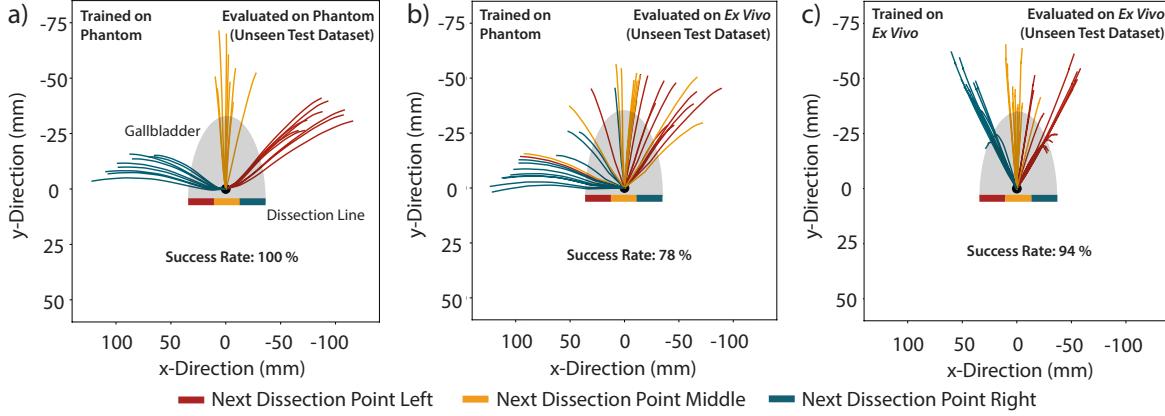


Fig. 4. Evaluation of DKMP on the phantom and *ex vivo* test datasets: a) Results of DKMP trained on phantom retractions and evaluated on the phantom test dataset. b) shows the results of the evaluation on the *ex vivo* test dataset. c) Results of DKMP trained on *ex vivo* retractions and evaluated on the *ex vivo* test dataset. The *dissection line* is subdivided into three areas where possible next *dissection points* can be and represented in their respective colors (left: red, middle: yellow, left: blue). The related retraction trajectories are represented in matching colors. View from above. The gallbladder is visualized in grey.

TABLE I
GALLBLADDER RETRACTION PREDICTION ACCURACY OF DKMP AND DEEPMP BASELINE ON PHANTOM AND *ex vivo* TEST DATASET.

Evaluated on	Trained on	Method	Endpoint and trajectory acc. $p \pm \sigma_p / \tau \pm \sigma_\tau$ (in mm)
Phantom	Phantom	DKMP	$17.7 \pm 11.8 / 12.0 \pm 6.3$
		DeepMP	$19.3 \pm 13.9 / 12.2 \pm 7.3$
<i>Ex Vivo</i>	Phantom	DKMP	$44.7 \pm 33.6 / 27.0 \pm 18.7$
		DeepMP	$58.4 \pm 23.6 / 35.9 \pm 12.6$
<i>Ex Vivo</i>	DKMP	DKMP	$12.2 \pm 10.8 / 8.4 \pm 5.6$
	DeepMP	DeepMP	$10.5 \pm 9.0 / 8.3 \pm 4.0$

The gallbladder retractions performed by the semi-autonomous assistance were evaluated by the clinical expert based on the correct retraction direction and force.

IV. RESULTS

A. Evaluation on Test Datasets

The test datasets include 27 unknown data samples for the phantom dataset and 36 unknown data samples for the *ex vivo* dataset, which are uniformly distributed over the variations of the gallbladder retractions. The results of the trajectory points and endpoint prediction accuracy of DKMP and DeepMP are depicted in Table I. DKMP reaches an endpoint accuracy of 17.7 ± 11.8 mm on the phantom dataset and 12.2 ± 10.8 mm on the *ex vivo* dataset. In comparison, DeepMP reaches an endpoint accuracy of 19.3 ± 13.9 mm on the phantom dataset and 10.5 ± 9.0 mm on the *ex vivo* dataset. When evaluating the phantom-trained methods on the *ex vivo* dataset, DKMP and DeepMP show a much larger prediction error of 44.7 ± 33.6 mm and 58.4 ± 23.6 mm respectively. The accuracies $\tau \pm \sigma_\tau$ of the trajectory points are stated in Table I. Overall, DKMP and the DeepMP baseline show a comparable prediction accuracy on the phantom and *ex vivo* dataset.

The phantom-trained DKMP predict the retraction directions when evaluated on the phantom test dataset (Figure 4a) with a success rate of 100%. When evaluated on the *ex vivo*

dataset, the phantom-trained DKMP partially predict wrong retraction directions, which is aligned with the observed higher prediction error (Figure 4b) reaching a success rate of only 78%. The *ex vivo*-trained DKMP evaluated on the *ex vivo* dataset predicts the correct retraction directions (Figure 4c) with a success rate of 94%. Overall, the evaluation of DKMP on the test datasets shows accurate prediction of the retraction directions.

B. Evaluation in Phantom and *Ex Vivo* Surgeries

A total of 35 gallbladder retractions were performed during five phantom surgeries (six with the *next dissection point* being on the left of the dissection line, 17 in the middle, and 12 on the right) as depicted in figure 5a. The average time spent for selecting the surgical landmarks in the graphical user interface was 44 seconds, while the average time for predicting the gallbladder retraction trajectory was 0.0034 seconds (on CPU). In total, 32 out of the 35 gallbladder retraction trajectories during the phantom surgeries were rated as successful by the clinical expert, resulting in a success rate of 91%. During the five *ex vivo* porcine liver surgeries, a total of 52 gallbladder retractions were performed (Figure 5b). Among these gallbladder retractions, 17 18 retractions had the *next dissection point* on the left of the dissection line, 19 in the middle, and 16 on the right. The average time spent for selecting the surgical landmarks in the graphical user interface was 64 seconds, while the average time for predicting the gallbladder retraction trajectory was 0.0053 seconds (on CPU). The average total procedure time across the five experiments was 41.8 ± 21.8 minutes. The procedure times decreased from 67.2 and 59.0 minutes in the first two experiments to 38.2, 32.1, and 12.5 minutes in the final three experiments. In total, 48 out of the 52 gallbladder retraction trajectories during the *ex vivo* porcine liver surgery were rated as successful by the medical expert, resulting in a success rate of 92%. The semi-autonomous robotic assistance using DKMP resulted in success rates of 91% and 92% during phantom and *ex*

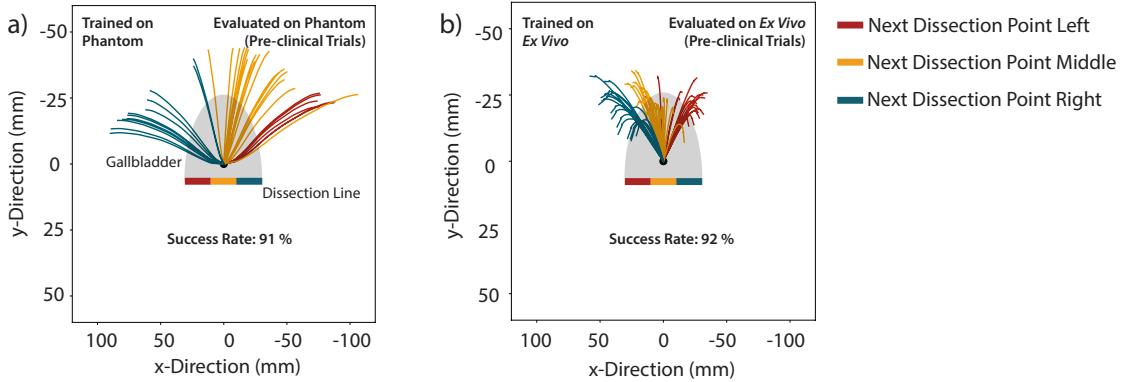


Fig. 5. Evaluation of DKMP during phantom and *ex vivo* surgeries in pre-clinical trials by medical expert: Gallbladder retraction trajectories with different *next dissection points* on the dissection line defined by the clinical expert, during phantom surgeries (a) and *ex vivo* porcine liver surgeries (b).

vivo surgeries. During the pre-clinical evaluation on unseen phantoms and *ex vivo* organs the gallbladder was always pulled in relation to the retraction direction until the force limit was reached, resulting in a favorable state for initiating dissection.

V. DISCUSSION

Here, we propose domain knowledge-based movement primitives (DKMP) to provide a semi-autonomous robotic system for assistance during cholecystectomies and with the goal of improving the interpretability of semi-autonomous robotic assistance systems for surgeons, addressing the shortage of clinical personnel and regulatory concerns inherent to surgical applications.

When evaluated on the test datasets, we achieve performance comparable to the DeepMP baseline. However, a key distinction lies in the input features used by DKMP, which are based on surgeon defined landmarks and calculated 3D point cloud feature vectors of the surgical scene. Thus, they are easier to interpret compared to the learned features of DeepMP. This improvement enhances the understanding and trustworthiness of the system for surgeons. Furthermore, our method benefits from ProMP conditioning, resulting in smoother trajectories compared to the learning of complete ProMP trajectories used in DeepMP. DKMP accurately predicts the retraction directions when trained and evaluated on the same scenario (phantom or *ex vivo*), achieving success rates of 100% and 94%, respectively. However, when DKMP was trained solely on phantom data and applied to *ex vivo* organs the success rate was only 78%. This indicates that the phantom and *ex vivo* organs differ in their biomedical properties, so that a training on real tissue is necessary to achieve the desired behavior. Moreover, the trajectories on the phantom are more spread out compared to the ones on the *ex vivo* porcine livers. These observations can be explained with the difference in material properties of the phantom and *ex vivo* porcine livers, as presented in prior work [24]. The development of phantoms with more realistic material properties could close this gap between the phantom and *ex vivo* tissue.

The potential of DKMP was further demonstrated in phantom and *ex vivo* surgeries, where we achieve a rate of 91% and 92% of successful gallbladder retractions. Notably, the method demonstrated effective performance with real *ex vivo* tissue, further supporting its applicability in clinical settings. For the trajectories during *ex vivo* surgeries, our method learned a combination of stretching the gallbladder in the right direction and subsequently following the gallbladder's natural curvature, as depicted in Figure 5b. This indicates that retraction based on single point prediction is insufficient and highlights the advantages of our trajectory generation. These findings highlight the potential of DKMP as a reliable and interpretable tool for robotic surgical assistance, addressing both performance and interpretability requirements. In comparison to the test dataset evaluation, the retraction trajectories of the pre-clinical trials are more spread out, which can be explained with larger variations of *next dissection points* used by the medical expert during surgeries compared to demonstrations in the dataset conducted by the technical research staff. The medical expert used *next dissection points*, distributed along the entire *dissection line*. To present our results more easily, all *next dissection points* were categorized in one of the three classes (left, middle, right).

Despite the promising results, DKMP has several limitations that will be addressed in future work. Currently, the detection of surgical landmarks relies on the surgeon identifying them through a graphical user interface (GUI). The focus of this work was to solve trajectory generation and the pre-clinical implementation of a robotic assistance system. As works from the field have already proven the feasibility of feature detection and segmentation of point clouds [29], [30], future developments should focus on integrating robust algorithms for the automatic detection of surgical landmarks. Additionally, although the input and output of the neural network-based movement prediction in DKMP can be interpreted, the decision-making process within the neural network remains a black box. This lack of transparency in the prediction mechanism poses challenges for understanding and validating the system's behavior under diverse surgical conditions.

While the evaluation of the semi-autonomous robotic as-

sistance system in phantom and *ex vivo* surgeries shows promising results, there are still several challenges towards clinical adoption. The use of a conventional stereo camera restricted the experimental setup to mimicking minimally invasive surgery by performing "open" surgery with laparoscopic instruments (see Figure 1). With a 3D endoscope, the experimental setup can be adapted for laparoscopic technique with laparoscopic instruments and with the upper face of the liver facing downwards, which more faithfully depicts the *in vivo* scenario with variable lighting conditions and robot motion constraints. Finally, although DKMP have been shown to perform well with a limited amount of data, collecting sufficient and diverse datasets for *in vivo* scenarios remains a significant challenge. Addressing these limitations is critical for advancing the system towards clinical adoption and enhancing its reliability in real-world surgical applications. The proposed DKMP framework presents several opportunities for future research and development in robotic surgery assistance.

VI. CONCLUSION

This study introduces a semi-autonomous robotic assistance system for gallbladder retraction during minimally invasive cholecystectomy. DKMP generate gallbladder retraction motions by combining 3D point cloud feature vectors based on domain knowledge with neural network-based movement prediction and ProMP for trajectory generation. The experiments show that DKMP successfully learn gallbladder retraction on phantoms and *ex vivo* porcine livers. While resulting in comparable success rates, DKMP improve the surgeon's ability to interpret the robot behavior through the usage of domain knowledge compared to end-to-end imitation learning methods, such as DeepMP. In summary, DKMP offer a data-efficient and interpretable method for learning complex surgical skills, such as gallbladder retraction. **Its ability to interpret the movement predictions input and output through the usage of domain knowledge make DKMP a promising approach for safe (semi-)autonomous assistance in robotic surgery.** Future works plan to automate the extraction of surgical landmarks to enable a simplified user interface for surgeons, as well as the extension of the system to provide assistance in additional phases of cholecystectomy as well as other procedures.

REFERENCES

- [1] L. Maier-Hein, *et al.*, "Surgical data science for next-generation interventions," *Nature Biomedical Engineering*, vol. 1, pp. 691–696, 9 2017.
- [2] R. Taylor, N. Simaan, A. Menciassi, and G. Yang, "Surgical robotics and computer-integrated interventional medicine," *Proceedings of the IEEE*, vol. 110, 2022.
- [3] P. Fiorini, K. Y. Goldberg, Y. Liu, and R. H. Taylor, "Concepts and trends in autonomy for robot-assisted surgery," *Proceedings of the IEEE*, vol. 110, 2022.
- [4] T. Haidegger, "Autonomy for surgical robots: Concepts and paradigms," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, pp. 65–76, 4 2019.
- [5] G. Z. Yang, *et al.*, "Medical robotics-regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, 3 2017.
- [6] P. M. Scheikl, *et al.*, "Cooperative assistance in robotic surgery through multi-agent reinforcement learning," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.
- [7] R. Zhang, *et al.*, "A step towards conditional autonomy - robotic appendectomy," *IEEE Robotics and Automation Letters*, vol. 8, 2023.
- [8] V. M. Varier, D. K. Rajamani, N. Goldfarb, F. Tavakkolmoghaddam, A. Munawar, and G. S. Fischer, "Collaborative suturing: A reinforcement learning approach to automate hand-off task in suturing for surgical robots," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020.
- [9] H. Saeidi, *et al.*, "Autonomous robotic laparoscopic surgery for intestinal anastomosis," *Science Robotics*, vol. 7, p. 2908, 2022.
- [10] M. Wagner, *et al.*, "A learning robot for cognitive camera control in minimally invasive surgery," *Surgical Endoscopy*, vol. 35, no. 9, pp. 5365–5374, 2021.
- [11] A. Attanasio, *et al.*, "Autonomous tissue retraction in robotic assisted minimally invasive surgery - a feasibility study," *IEEE Robotics and Automation Letters*, vol. 5, 2020.
- [12] M. Wagner, *et al.*, "Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the heichole benchmark," *Medical Image Analysis*, vol. 86, 5 2023.
- [13] C. Shin, P. W. Ferguson, S. A. Pedram, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous tissue manipulation via surgical robot using learning based model predictive control," *2019 International Conference on Robotics and Automation (ICRA)*, 2019.
- [14] N. D. Nguyen, T. Nguyen, S. Nahavandi, A. Bhatti, and G. Guest, "Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery," *2019 IEEE International Systems Conference (SysCon)*, pp. 1–7, 2019.
- [15] X. Liang, C.-P. Wang, N. U. Shinde, F. Liu, F. Richter, and M. Yip, "Medic: Autonomous surgical robotic assistance to maximizing exposure for dissection and cauterization," *arXiv preprint arXiv:2409.14287*, 2024.
- [16] K.-H. Oh, L. Borgioli, M. Žefran, V. Valle, and P. C. Giulianotti, "Autonomous dissection in robotic cholecystectomy," *arXiv preprint arXiv:2503.00666*, 2025.
- [17] A. Pore, E. Tagliabue, M. Piccinelli, D. Dall'Alba, A. Casals, and P. Fiorini, "Learning from demonstrations for autonomous soft-tissue retraction," *2021 International Symposium on Medical Robotics (ISMР)*, 10 2021.
- [18] K. L. Schwander, D. Dall'alba, P. T. Jensen, P. Fiorini, and T. R. Savarimuthu, "Autonomous needle manipulation for robotic surgical suturing based on skills learned from demonstration," *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 2021.
- [19] H. Su, A. Mariani, S. E. Ovur, A. Menciassi, G. Ferrigno, and E. D. Momi, "Toward teaching by demonstration for robot-assisted minimally invasive surgery," *IEEE Transactions on Automation Science and Engineering*, vol. 18, pp. 484–494, 2021.
- [20] P. M. Scheikl, *et al.*, "Movement primitive diffusion: Learning gentle robotic manipulation of deformable objects," *IEEE Robotics and Automation Letters*, vol. 9, pp. 5338–5345, 2024.
- [21] O. Sanni, G. Bonvicini, M. A. Khan, P. C. López-Custodio, K. Nazari, and A. M. Ghalamzan, "Deep movement primitives: Toward breast cancer examination robot," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 12 126–12 134, 2022.
- [22] D. Meli, E. Tagliabue, D. Dall'alba, and P. Fiorini, "Autonomous tissue retraction with a biomechanically informed logic based framework," *2021 IEEE International Symposium on Medical Robotics (ISMР)*, 2021.
- [23] A. Soleymani, X. Li, M. Tavakoli, and S. Member, "A domain-adapted machine learning approach for visual evaluation and interpretation of robot-assisted surgery skills," *IEEE Robotics and Automation Letters*, vol. 7, 2022.
- [24] A. Schuessler, R. Younis, J. Paik, M. Wagner, F. Mathis-Ullrich, and C. Kunz, "Interactive surgical liver phantom for cholecystectomy training," *Current Directions in Biomedical Engineering*, vol. 10, pp. 95–98, 2024.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition; deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [26] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Probabilistic movement primitives," *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013.
- [27] P. Alexandros, D. Christian, P. Jan, and N. Gerhard, "Using probabilistic movement primitives in robotics," *Autonomous Robots*, vol. 42, pp. 529–551, 2018.
- [28] A. Lazaric and M. Ghavamzadeh, "Bayesian multi-task reinforcement learning," *ICML - 27th International Conference on Machine Learning*, 2010.
- [29] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 652–660, 2017.

- [30] B. Alt, *et al.*, “Lapseg3d: Weakly supervised semantic segmentation of point clouds representing laparoscopic scenes,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5265–5270, 2022.