

Learning from imperfect demonstrations in a surgical training task

Yi Hu^a, Yafei Ou^a, August Sieben^b, Zahra Samadikhoshkho^b, Bin Zheng^c, Jun Jin^{a,e}, Mahdi Tavakoli^{a,d}

^aDepartment of Electrical and Computer Engineering, University of Alberta, Edmonton, T6G 2R3, Canada

^bDepartment of Mechanical Engineering, University of Alberta, Edmonton, T6G 2R3, Canada

^cDepartment of Surgery, University of Alberta, Edmonton, T6G 2R3, Canada

^dDepartment of Biomedical Engineering, University of Alberta, Edmonton, T6G 2R3, Canada

^eAlberta Machine Intelligence Institute (AMII), Edmonton, T5J 1S5, Canada

Abstract

Robotic surgery offers several advantages over traditional techniques, including improved precision, greater consistency, and enhanced dexterity. Learning from demonstrations (LfD) is a promising approach for transferring expert skills to robots, thereby alleviating clinicians' physical workload. However, a major challenge in surgical robotics is that demonstration data often includes suboptimal or failed behaviors due to human error, fatigue, or the inherent complexity of surgical tasks. Discarding such imperfect data results in the loss of valuable information and hinders the scalability of data-driven surgical skill acquisition. In this work, we propose a novel LfD optimization framework capable of learning from a broad spectrum of demonstrations—including successful, suboptimal, and failed attempts. Our method employs a dual probabilistic modeling strategy to encode demonstrations and formulates a multi-objective optimization problem under novel problem conditions to find an optimal reproduction. We validate our approach on the standard ring-and-rail task, a representative surgical training task requiring high-precision and dexterous manipulation. Real-world experiments using the da Vinci Research Kit (dVRK) show that, even in the presence of failure cases within the demonstration set, our method produces optimized trajectories that enable the patient-side manipulator to successfully guide the ring along the curved wire without contact. These results demonstrate the robustness and effectiveness of our approach in learning from imperfect data, underscoring its potential for real-world deployment in robot-assisted surgery.

Keywords: Robotic surgical task, Learning from demonstration, Autonomous surgical task

1. Introduction

Robotic surgery holds significant promise for improving surgical precision, consistency, and enabling minimally invasive interventions. Despite this potential, the automation of surgical tasks remains a formidable challenge due to their safety-critical nature, high dexterity requirements, and operation within constrained anatomical environments. Unlike industrial tasks, surgical subtasks—such as suturing [1], retraction [2], and microsurgery—require expert-level coordination and fine-grained motion control [3]. Encoding such skills through hand-crafted control policies is extremely difficult, as it requires anticipating a vast number of edge cases. Likewise, trial-and-error learning approaches are generally infeasible due to the unacceptable risks of failure and damage to tissue during the learning process.

Learning from demonstration (LfD) has emerged as a compelling paradigm for automating complex surgical tasks by leveraging the expertise of skilled human operators. By allowing robots to imitate human-provided trajectories, LfD circumvents the need for manual programming or unsafe exploration, and instead leverages the expertise embedded in surgical demonstrations. Prior work has applied LfD to various surgical tasks, including suturing [4, 5], ultrasound scanning [6, 7], retraction [8], tool navigation [9], and laparoscopic camera positioning [10]. These studies typically assume access to

high-quality, flawless demonstrations to ensure reliable policy learning.



Figure 1: Ring-and-rail task for training/assessing fundamental surgical skill. The operator teleoperates the PSM to guide a rigid ring along a narrow, curved metal wire without making contact. This surgical training task simulates critical aspects of real surgical procedures, such as hand-eye coordination, fine motor control, spatial planning, and operation within constrained environments.

However, in real-world surgical environments, collecting consistent, high-quality demonstrations is non-trivial [11]. Even expert surgeons are susceptible to producing imperfect or failed demonstrations due to occlusions, anatomical variability, tool-tissue interaction complexity, or operator fatigue. As a re-

sult, demonstration datasets often contain a wide spectrum of trajectories, ranging from successful and near-optimal to completely failed attempts. While suboptimal behaviors are typically discarded in conventional LfD pipelines, they may contain critical information about failure modes, unsafe regions, or challenging aspects of the task [12]. This underscores the need for learning frameworks that can effectively leverage both successful and imperfect demonstrations to enhance the robustness and generalizability of autonomous surgical skill learning.

In this work, we investigate a fundamental question in robotic surgical task learning: whether a robot can learn an optimal execution policy from demonstration data that includes both successful and failed attempts. This question is central to enabling robust learning in realistic surgical settings, where high-quality, error-free demonstrations are often limited. To this end, we propose a novel LfD framework that explicitly models and learns from a mixture of successful, suboptimal, and failed demonstrations. Our approach leverages a hybrid probabilistic representation: Gaussian Mixture Models (GMMs) are used to capture the spatial-temporal patterns of successful behaviors, while Donut Mixture Models (DMMs) [13] are introduced to explore the potential expert behavior from failure-prone regions. By jointly modeling perfect and imperfect demonstrations, our framework learns a richer task representation and improves sample efficiency. We formulate trajectory generation as a multi-objective optimization problem that balances multiple criteria, including imitation fidelity, smoothness, task compliance, and failure avoidance. Unlike Trajectory Learning from Failed and Successful Demonstrations (TLFSD) [14] that subtract failed data or require confidence scoring, our method integrates the underlying structure of both successful and failed trajectories, enabling it to discover actionable insights from failure cases and fuse them constructively with successful data. Our key contributions are as follows:

- This work is the first attempt to address the challenge of learning from imperfect demonstrations in the ring-and-rail task designed for surgical skill training.
- The proposed framework can extract latent successful behaviors from imperfect demonstrations, significantly improving its practicality in real-world surgical learning scenarios where high-quality data is scarce.
- We propose a multi-objective optimization strategy that jointly balances imitation fidelity, task compliance, and failure avoidance to enable the acquisition of precise and robust surgical skills.

To evaluate our proposed framework, we select the ring-and-rail task, a widely adopted surgical skill training task to train or assess surgical motion skills for both novice and experienced surgeons. In this task, a surgeon is required to guide a ring along a narrow and curved metal wire without making any contact. The task emulates key aspects of real surgical procedures, including hand-eye coordination, fine motor control, spatial planning, and constrained motion.

In real-world validation, we conduct experiments using the da Vinci Research Kit (dVRK) [15], a widely adopted research platform in surgical robotics. Our experimental setup includes a Patient Side Manipulator (PSM) equipped with a grasper to manipulate the ring along the rail, simulating constrained surgical motions. To detect task failures, we integrate an LED-based contact sensing system that activates when the ring touches the wire, as shown in Fig. 1. These experiments not only demonstrate the practical viability of our method in real-world surgical contexts, but also highlight its potential to enhance surgical training and execution—particularly in high-precision domains such as microsurgery and neurosurgery.

2. Related Work

This work is inspired by the following topics from both the autonomous surgical robots research community and imitation learning from imperfect demonstrations.

Deep Learning for Surgical Task Modeling Deep learning has driven substantial advances in surgical scene understanding and control. Huang et al. [7] extracted attention regions from expert demonstrations for carotid ultrasound scanning to ensure smooth and robust transitions. Zhang et al. [16] developed a needle navigation system for retinal microsurgery using convolutional neural networks to predict distance vectors. Gonzalez et al. [17] proposed a data-efficient action recognition system for telesurgery, leveraging simulated data to reduce training requirements. Qin et al. [18] designed a deep learning framework to estimate both coarse and fine-grained surgical states during robotic hernia repair. While impactful, these works require large-scale, clean, annotated datasets and do not address policy learning from imperfect demonstrations. In contrast, our approach circumvents the need for network training and instead focuses on extracting optimal behaviors directly from mixed-quality demonstrations.

Reinforcement Learning for Surgical Automation Reinforcement learning (RL) has been employed to learn surgical policies through reward-guided exploration. Shahkoo et al. [19] used deep RL to optimize tissue tensioning during robotic cutting. Scheikl et al. [20] introduced an image-based RL framework for deformable tissue manipulation. Ou et al. [21] explored deep RL for blood suction automation. Long et al. [22] studied human-in-the-loop RL as a hybrid learning paradigm for surgical systems, and Bi et al. [23] proposed a simulation-based RL framework for navigating ultrasound probes to standard anatomical views. Despite their promise, RL approaches face challenges in surgical contexts due to the difficulty of designing suitable reward functions and the high sample complexity of iterative policy refinement. Our framework mitigates these issues by avoiding reward engineering and relying instead on information-rich demonstrations.

Movement Primitives in Surgical Skill Learning Movement primitives (MPs) have long been a foundational approach in surgical automation due to their favorable properties in terms of safety, sample efficiency, and ease of generalization. Scheikl et al. [24] combined diffusion-based imitation learning with probabilistic dynamic movement primitives (DMPs) to handle

deformable object manipulation. Su et al. [25] applied GMM-based DMPs in the context of robot-assisted minimally invasive surgery, enabling the encoding of smooth and adaptable motion trajectories. Schwaner et al. [4] used DMPs to model low-level surgical action primitives and compose more complex behaviors. Pan et al. [26] utilized discrete DMPs to capture the spatial-temporal structure of palm trajectories during teleoperated control. However, traditional MP-based approaches are sensitive to distribution shifts and compounding errors, particularly under noisy or inconsistent demonstrations. In contrast, our method explicitly models both successful and failed demonstrations, enabling robust policy learning even when the dataset consists entirely of failed attempts.

Imitation Learning from Imperfect Demonstrations Recent studies have explored learning from demonstrations of varying quality. Cao et al. [27] introduced a hybrid metric combining feasibility and optimality to assess the usefulness of suboptimal trajectories. Chou et al. [28] learned logical task structures from imperfect demonstrations using linear temporal logic. Zhu et al. [29] proposed a ranking-based imitation framework suitable for settings with sparse or delayed rewards. Donut distributions [13, 30] and subsequent optimization-based methods [14] have shown potential in utilizing suboptimal data by explicitly modeling negative regions. However, many of these approaches either depend on manually assigned confidence scores or require a significant proportion of expert demonstrations. Our method departs from these paradigms by avoiding reliance on confidence estimation and instead leveraging the internal structure of both successful and failed demonstrations through probabilistic modeling.

3. Methodology

In this section, we present our proposed framework for learning from imperfect demonstrations. The methodology comprises three key components: (i) demonstration collection; (ii) modeling successful and failed demonstrations; and (iii) generation of optimized reproductions via a multi-objective trajectory optimization process.

3.1. Demonstration Collection

Given a set of mixed surgical task demonstration $\mathcal{D}$ that contain both successful, suboptimal and failed executions. The dataset is partitioned into two subsets, $\mathcal{D}_s$ and $\mathcal{D}_f$, corresponding to successful and failed demonstrations, respectively. Formally, we define:

$$\mathcal{D} = \{\mathcal{D}_s, \mathcal{D}_f\}, \quad \mathcal{D}_s = \{X_s^1, \dots, X_s^M\}, \quad \mathcal{D}_f = \{X_f^1, \dots, X_f^N\} \quad (1)$$

where X_s^m and X_f^n denote successful and failed trajectories, and M, N are the number of samples. Each trajectory X^j is a discrete, finite-length sequence represented in Cartesian task space:

$$X^j = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]^\top \in \mathbb{R}^{T \times Q} \quad (2)$$

where $\mathbf{x}_t = [x_{t(1)}, x_{t(2)}, \dots, x_{t(Q)}] \in \mathbb{R}^Q$ denotes the Q -dimensional observation at time step t , and T is the trajectory length.

To account for variations in execution timing across demonstrations, we apply Dynamic Time Warping (DTW) to temporally align the raw trajectories. The aligned trajectories are then uniformly resampled for consistent modeling. When dealing with a mixed set of demonstrations that includes both successful and failed trajectories, TLFSD [14] proposes to avoid failure by subtracting the distribution of failed behaviors from that of successful ones. In contrast, our approach aims to extract and preserve the informative motion patterns embedded within failed demonstrations, enhancing them through integration with successful trajectories rather than discarding them outright.

3.2. Demonstration Modeling

To effectively learn from demonstrations of varying quality—including successful, failed, and intermediate cases—we adopt a dual probabilistic modeling approach. This section outlines the formulation used to model and learn from such mixed-quality demonstration datasets.

3.2.1. Modeling Successful Demonstrations with Gaussian Mixture Models

Let the set of successful demonstrations $\mathcal{D}_s = \{X_s^1, \dots, X_s^M\}$ consist of M trajectories, where each trajectory $X_s^m \in \mathbb{R}^{T \times Q}$ is a sequence of T time-aligned Cartesian observations in Q -dimensional space. To model the distribution of these trajectories, we employ GMMs that captures the joint probability density over time t and observation $\mathbf{x} \in \mathbb{R}^Q$ [31]:

$$p(t, \mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(t, \mathbf{x} \mid \mu_k, \Sigma_k), \quad (3)$$

where K is the number of Gaussian components, π_k is the prior probability of the k^{th} component, and $\mathcal{N}(t, \mathbf{x} \mid \mu_k, \Sigma_k)$ is the multivariate Gaussian distribution with mean $\mu_k = [\mu_t^k, \mu_x^k]^T$ and covariance matrix $\Sigma_k = \begin{bmatrix} \Sigma_{tt}^k & \Sigma_{tx}^k \\ \Sigma_{xt}^k & \Sigma_{xx}^k \end{bmatrix}$.

The complete set of GMM parameters is denoted as $\theta = \{K, \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K\}$. These parameters are estimated using the Expectation-Maximization (EM) algorithm to maximize the likelihood of the observed successful trajectories. The optimal number of components K is determined using the Bayesian Information Criterion (BIC) to balance model fit and complexity.

After training the GMM, we apply Gaussian Mixture Regression (GMR) to estimate the conditional distribution of $\mathbf{x}$ given time t . The conditional distribution is defined as:

$$p(\mathbf{x} \mid t) = \mathcal{N}(\mathbf{x}; \hat{\mu}_x(t), \hat{\Sigma}_x(t)), \quad (4)$$

where the conditional mean and covariance are computed as: $\hat{\mu}_x(t) = \sum_{k=1}^K \pi_k (\mu_x^k + \Sigma_{xt}^k (\Sigma_{tt}^k)^{-1} (t - \mu_t^k))$, $\hat{\Sigma}_x(t) = \sum_{k=1}^K \pi_k (\Sigma_{xx}^k - \Sigma_{xt}^k (\Sigma_{tt}^k)^{-1} \Sigma_{tx}^k)$. The optimal reproduction trajectory is given by the sequence of predicted means, i.e., $X_{\text{GMM}} = \hat{\mu}_x(t)$.

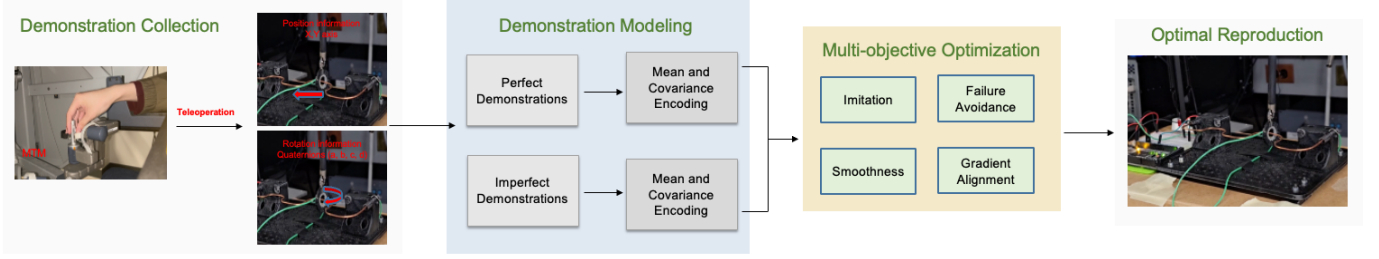


Figure 2: Overview of the proposed framework for surgical skill reproduction in the surgical training task. Demonstrations are collected via teleoperation using the MTM to control the PSM of the dVRK. Both perfect and imperfect executions are modeled through probabilistic encodings and used in a multi-objective optimization process. The optimized trajectory is then autonomously executed by the PSM.

This approach yields smooth and consistent trajectories that closely follow the statistical tendencies of expert demonstrations. However, GMMs trained solely on successful data cannot capture or correct for suboptimal behaviors when trained on mixed-quality data. This limitation motivates the integration of failure modeling, as addressed in the subsequent section.

3.2.2. Modeling Failed Demonstrations with Donut Mixture Models

Let the set of failed demonstrations $\mathcal{D}_f = \{X_f^1, \dots, X_f^N\}$ consist of N trajectories, where each trajectory $X_f^n \in \mathbb{R}^{T \times Q}$ is a sequence of T time-aligned Cartesian observations in Q -dimensional space. To model the structure of these failed executions and uncover latent successful behavior, we employ DMMs, which provide a principled approach to probabilistically represent failure-prone data while retaining informative motion cues.

DMMs, introduced by Grollman et al. [13], are designed to model deviations from expert-like behavior in failed demonstrations. Unlike GMMs, which model successful data directly, the DMM leverages the Donut distribution—a distribution constructed as the difference between two Gaussians sharing the same mean but differing in covariance:

$$D(\mathbf{x} | \mu_\alpha, \mu_\beta, \Sigma_\alpha, \Sigma_\beta, \gamma) = \gamma \mathcal{N}(\mathbf{x} | \mu_\alpha, \Sigma_\alpha) - (\gamma - 1) \mathcal{N}(\mathbf{x} | \mu_\beta, \Sigma_\beta), \quad (5)$$

where $\gamma > 1$ controls the relative weighting between the two Gaussian components. To simplify the formulation, we assume both components share the same mean, i.e., $\mu_\alpha = \mu_\beta = \mu$. The corresponding covariance matrices are scaled versions of a shared base covariance Σ , defined as:

$$\Sigma_\alpha = \frac{1}{r_\alpha^2} \Sigma, \quad \Sigma_\beta = \frac{1}{r_\beta^2} \Sigma,$$

where r_α and r_β are scalar spread parameters that control the “thickness” of the Donut distribution. Substituting these definitions, the Donut distribution becomes:

$$D(\mathbf{x} | \mu, \Sigma, r_\alpha, r_\beta, \gamma) = \gamma \mathcal{N}(\mathbf{x} | \mu, \Sigma/r_\alpha^2) - (\gamma - 1) \mathcal{N}(\mathbf{x} | \mu, \Sigma/r_\beta^2). \quad (6)$$

To control the degree of exploration during inference, we introduce an exploration parameter $\epsilon \in [0, 1]$, which modulates the spread of the distribution. When $\epsilon = 0$, the Donut distribution approximates a standard Gaussian. Conversely, when $\epsilon = 1$, the distribution encourages maximal deviation from the observed failures. The spread parameters r_α and r_β are defined as functions of ϵ :

$$r_o(\epsilon) = (1 - \epsilon) \cdot (r_o(0) - r_o(1)) + r_o(1), \quad (7)$$

where $r_o \in \{r_\alpha, r_\beta\}$. This formulation allows for smooth transitions between conservative and exploratory trajectory representations.

The full Donut Mixture Model is constructed analogously to the GMM:

$$p_{\text{DMM}}(\mathbf{x}) = \sum_{k=1}^K \hat{\pi}_k D(\mathbf{x} | \hat{\mu}_k, \hat{\Sigma}_k, \epsilon), \quad (8)$$

where $\hat{\pi}_k$, $\hat{\mu}_k$, and $\hat{\Sigma}_k$ denote the mixture weight, mean, and base covariance matrix of the k^{th} Donut component.

However, obtaining a closed-form analytical solution for the maximum of the DMM is generally intractable due to the non-convex nature of the distribution. To address this, the gradient ascent strategy is employed to iteratively converge toward a local maximum within the parameter space. The gradient of the DMM with respect to the input is computed as the weighted sum of the gradients of its constituent components and is given by:

$$\begin{aligned} \nabla_x D(\mathbf{x} | \mu, \Sigma, r_\alpha, r_\beta, \gamma) = & \mathcal{D}(\mathbf{x} | \mu, \Sigma, r_\alpha) \Sigma_\alpha^{-1} (\mathbf{x} - \mu) + \\ & (\gamma - 1) \mathcal{D}(\mathbf{x} | \mu, \Sigma, r_\beta) \Sigma_\beta^{-1} (\mathbf{x} - \mu) \end{aligned} \quad (9)$$

The objective of the DMM is to produce locally valid trajectories that intentionally diverge from regions associated with failure. Since failed demonstrations highlight areas to avoid, identifying a local optimum outside these regions is sufficient to guide the generation of safe and reliable trajectories.

In this work, we leverage the exploration parameter ϵ to effectively control the distance between the synthesized reproduction and the original failed data, enabling recovery of task-relevant motion patterns otherwise overlooked in traditional imitation frameworks.

3.2.3. Multi-Objective Optimization

Let the set of mixed surgical task demonstration $\mathcal{D} = \{\mathcal{D}_s, \mathcal{D}_f\}$ contain both perfect and imperfect demonstrations. Based on the learned models (GMMs for $\mathcal{D}_s$, DMMs for $\mathcal{D}_f$), we seek to synthesize a trajectory $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$, where each $\mathbf{X}_t \in \mathbb{R}^Q$, that is safe, executable, and task-compliant.

To this end, we formulate trajectory reproduction as a multi-objective optimization problem. The objective function incorporates constraints and priors from both successful and failed demonstrations, along with motion regularity and dynamic behavior alignment. The trajectory optimization problem is defined as:

$$J(\mathbf{X}) = J^{\text{imit}}(\mathbf{X}) + J^{\text{avoid}}(\mathbf{X}) + J^{\text{smooth}}(\mathbf{X}) + J^{\text{grad}}(\mathbf{X}), \quad (10)$$

where each cost term contributes to a specific aspect of motion generation, as described below.

Imitation Cost $J^{\text{imit}}(\mathbf{X})$: This term encourages adherence to expert-like behavior while avoiding failure-indicative motion. Let $\mathbf{m}_s(t)$ and $\mathbf{m}_f(t)$ denote the time-dependent mean trajectories derived from GMM and DMM models, respectively. The imitation cost is defined as:

$$J^{\text{imit}}(\mathbf{X}) = \alpha_1 \sum_{t=1}^T \|\mathbf{X}_t - \mathbf{m}_s(t)\|^2 + \alpha_2 \sum_{t=1}^T \|\mathbf{X}_t - \mathbf{m}_f(t)\|^2, \quad (11)$$

where α_1 and α_2 are positive scalar weights balancing attraction to expert data and repulsion from failure modes.

Failure Avoidance Cost $J^{\text{avoid}}(\mathbf{X})$: To promote safety, we penalize proximity to observed failure points $\mathcal{F} = \{\mathbf{x}_{t_f}^{(n)}\}$, derived from $\mathcal{D}_f$. A safety margin $d_{\max}$ ensures sufficient clearance:

$$J^{\text{avoid}}(\mathbf{X}) = \alpha_3 \sum_{t=1}^T \sum_{\mathbf{x}_{t_f}^{(n)} \in \mathcal{F}} \left(\|\mathbf{X}_t - \mathbf{x}_{t_f}^{(n)}\| - d_{\max} \right)^2, \quad (12)$$

where α_3 is a penalty coefficient that enforces a soft constraint against failure-prone regions. Defective segments, such as obstacle contacts or failure, are explicitly handled through the failure avoidance cost, which penalizes proximity to states observed in failed demonstrations. The entire demonstration trajectory is processed, rather than isolating only the defective segments. This preserves temporal continuity and allows the optimization to incorporate both successful and failure-prone regions, ultimately encouraging safe, coherent, and task-compliant trajectory reproduction.

Smoothness Cost $J^{\text{smooth}}(\mathbf{X})$: We incorporate a second-order difference penalty to ensure smooth and physically feasible motions. This term regularizes trajectory accelerations:

$$J^{\text{smooth}}(\mathbf{X}) = \alpha_4 \sum_{t=2}^{T-1} \|\mathbf{X}_{t+1} - 2\mathbf{X}_t + \mathbf{X}_{t-1}\|^2, \quad (13)$$

where α_4 controls the degree of motion regularization.

Gradient Alignment Cost $J^{\text{grad}}(\mathbf{X})$: To match not only the trajectory shape but also its temporal evolution, we align directional motion patterns. Let $\{\mathbf{x}_t^{(n)}\}_{n=1}^{N_s}$ denote time-aligned successful demonstrations. The gradient alignment cost is:

$$J^{\text{grad}}(\mathbf{X}) = \alpha_5 \sum_{n=1}^{N_s} \sum_{t=1}^{T-1} \left\| (\mathbf{X}_{t+1} - \mathbf{X}_t) - (\mathbf{x}_{t+1}^{(n)} - \mathbf{x}_t^{(n)}) \right\|^2, \quad (14)$$

where α_5 is a positive scalar weight promoting similar motion flow.

Optimization Procedure: The final trajectory $\mathbf{X}^*$ is obtained by solving the unconstrained optimization problem:

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} J(\mathbf{X}). \quad (15)$$

This formulation enables robust trajectory synthesis by jointly minimizing task deviation, failure risk, motion discontinuity, and dynamic inconsistency. The result is a safe, smooth, and task-compliant trajectory, even in the presence of noisy or imperfect demonstrations. The overall framework of the proposed method is shown in Fig. 2.

4. Experiments and Results

In this section, we validate the feasibility and effectiveness of the proposed framework through both simulation studies and real-world experiments. The simulation evaluations are designed to assess and compare the accuracy of the reproduced trajectories, while the real-world ring-and-rail experiments demonstrate the practical applicability of our approach in a representative surgical task setting.

4.1. Simulation Validation

To evaluate the effectiveness of the proposed framework in handling imperfect demonstrations, we conducted simulation experiments involving two representative trajectory patterns: sinusoidal and cosinusoidal. Each scenario consists of six synthetic demonstration trajectories, a subset of which intentionally contains failure-prone segments to simulate suboptimal executions. These simulations provide a controlled environment to assess the robustness of different learning-from-demonstration methods under mixed-quality data conditions.

Figure 3 presents the reproduction results obtained using three approaches: (1) GMMs baseline trained on all demonstrations without distinguishing failures, (2) TLFS [14], which attenuates failure influence by subtracting their distribution from that of successful demonstrations, and (3) the proposed method, which models successful and failed demonstrations separately via GMM and DMM, and combines them through multi-objective optimization.

In the sine trajectory task (Fig. 3a), the GMM-based reproduction closely follows failure-affected trajectories and fails to remain within the safety threshold. TLFS partially reduces the impact of failures but still produces segments that violate the safety margin. In contrast, the proposed method maintains

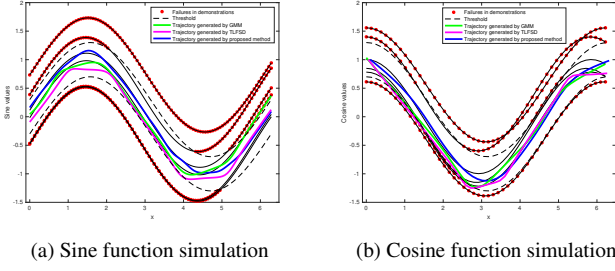


Figure 3: Trajectory reproduction results using GMM, TLFSD, and the proposed method. Dashed lines indicate failure thresholds. The proposed method consistently remains within the safe region and more accurately approximates the reference trajectory.

the entire trajectory within the safe region, effectively filtering out failure-related deviations. Similar trends are observed in the cosine trajectory task (Fig. 3b). The proposed method generates trajectories that are both more accurate and safety-compliant, whereas GMM and TLFSD exhibit either overshooting or underfitting behaviors in critical segments.

Quantitative evaluation is performed using Root Mean Square Error (RMSE) with respect to the reference (ground-truth) trajectory. The results are as follows: Sine Simulation: 0.021 for proposed method, 0.202 for GMM, 0.081 for TLFSD; Cosine Simulation: 0.123 for proposed method, 0.173 from GMM, 0.244 for TLFSD; These results indicate that the proposed method achieves the lowest reproduction error in both tasks. Compared to GMM and TLFSD, it improves accuracy by approximately 1.4–9.6 times, while maintaining adherence to safety constraints.

Overall, the simulation results support the central claim of this work: explicitly modeling both successful and failed demonstrations enables more accurate and robust trajectory reproduction. In contrast to existing methods that either ignore or heuristically down-weight failed demonstrations, the proposed approach leverages both sources of information through a principled optimization framework, resulting in improved performance under imperfect training conditions.

4.2. Real-world Experiment

To assess the practical applicability of the proposed framework in a real robotic surgical task, we conduct experiments using the dVRK on the ring-and-rail task. This task is widely used in surgical training to evaluate dexterity, trajectory accuracy, and motion control under constrained conditions for both novice and experienced surgeons. It presents challenges analogous to those encountered in surgical procedures, including spatial constraints and collision avoidance.

4.2.1. Experimental Setup

The experimental platform is designed to evaluate the effectiveness of the proposed framework in a realistic robotic surgical task that mimics surgical training. As shown in Fig. 4, the setup consists of a curved metallic wire rail mounted horizontally on a planar workspace. The task requires the robot

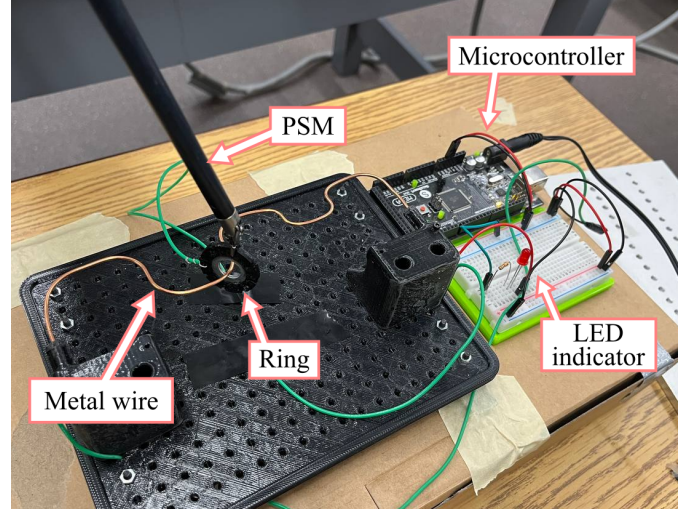
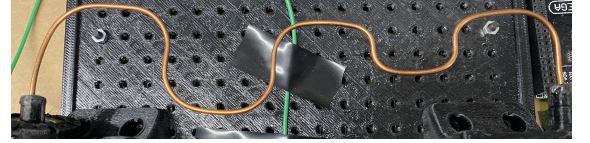


Figure 4: Ring-and-rail experimental platform used for evaluating learning from imperfect demonstrations. The dVRK PSM manipulates a rigid ring along a curved metallic wire while avoiding contact. The setup integrates an LED-based contact detection circuit, triggered by a microcontroller upon voltage drop when the ring touches the wire. This mechanism enables real-time failure detection and automatic labeling of demonstration quality.

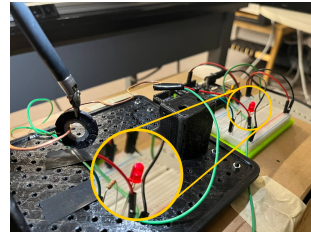


(a) Rail geometry for trajectory execution

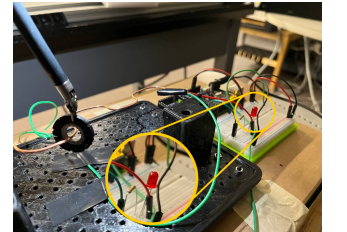


(b) Segmented view for illustration of curvature complexity

Figure 5: Curved wire geometry of the ring-and-rail task. The path includes regions of varying curvature and narrow passages, requiring high-precision motion.



(a) LED ON: contact detected



(b) LED OFF: no contact

Figure 6: Contact detection using an LED-based sensing circuit. A contact event between the ring and the wire triggers a voltage drop, activating the LED (left). In the absence of contact, the LED remains off (right). This binary signal enables automated annotation of failure segments during demonstration.

to guide a rigid ring along the rail while avoiding any physical contact. This setup simulates common challenges in surgical procedures, such as constrained motion, fine motor control, and contact avoidance.

The platform is built on the dVRK, comprising a Master Tool Manipulator (MTM) and a PSM. During teleoperation, the MTM is controlled by a human operator to collect demonstration trajectories. For autonomous execution, the learned policies are deployed on the PSM, which reproduces the desired motion without manual intervention.

To enable real-time detection of contact events, a microcontroller-based sensing system is integrated into the rail. When the ring touches the wire, a voltage drop occurs in the circuit, triggering a light-emitting diode (LED) as shown in Fig. 6. An overhead camera monitors the LED state throughout task execution, allowing frame-level annotation of contact events. This automatic labeling process supports consistent classification of demonstrations into successful and failed segments. The rail geometry is illustrated in Fig. 5 to illustrate the complexity of the ring-and-rail task. The wire follows a continuous, curved path with varying curvature and narrow clearances, which increases the difficulty of trajectory execution.

4.2.2. Demonstration Collection

To emulate variability and operational challenges commonly encountered in surgical settings, controlled disturbances were applied during the teleoperated demonstrations. Perturbation forces were introduced at the end-effector of the MTM to simulate conditions such as reduced visibility or transient tool instability. Five demonstrations were collected, each beginning from a consistent initial pose and following the ring-and-rail path. The PSM was controlled via the MTM during teleoperation. A contact detection system, based on an LED-triggered voltage sensor, was used to identify ring-wire contact events in real time. Binary labels were assigned to each time step according to the LED state. The resulting dataset contains a mixture of successful and failure-prone segments, reflecting realistic variability in surgical task execution.

4.2.3. Optimal Trajectory Reproduction

The reproduction results based on the collected demonstrations are presented in Figs. 7–11, illustrating both Cartesian and rotational motion characteristics of the executed trajectories.

Figure 7 displays the end-effector trajectory in the X – Y plane. The proposed method accurately follows the rail path while consistently maintaining a safe clearance from the wire. Notably, the trajectory remains fully within the failure-free region, demonstrating the framework’s effectiveness in extracting safe motion profiles from imperfect training data. In contrast, the GMM-based method exhibits noticeable drift and deviation, particularly in sections of high curvature. These deviations frequently lead to contact with the rail, as indicated by overlapping with failure regions. TLFSF improves over GMM by attenuating failure-prone deviations, but still results in trajectory overshoot near sharp bends, compromising task safety and completion.

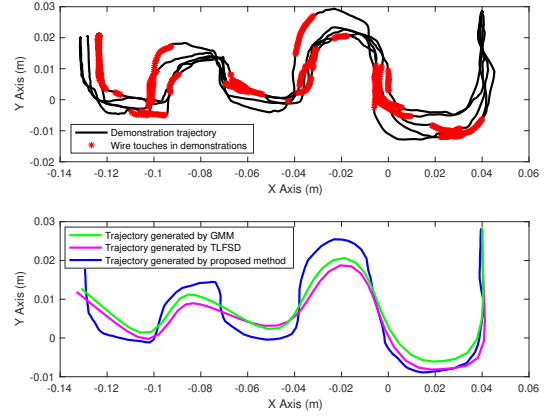


Figure 7: Reproduced trajectory in the X – Y plane. The proposed method closely follows the rail geometry while avoiding contact. Baseline methods (GMM, TLFSF) show drift and overlap with failure regions, particularly in curved segments.

Figure 8 shows the quaternion-based orientation of the end-effector across the trajectory. To further analyze rotational behavior, we decompose the orientation into roll, pitch, and yaw angles in Figs. 9–11. These results reveal clear differences between the methods. The proposed approach maintains smooth, continuous angular transitions across all three axes, indicative of stable and coordinated tool alignment. In contrast, both GMM and TLFSF exhibit sharp transitions and oscillations, particularly in the roll and yaw axes. Such abrupt angular changes can lead to tool misalignment and instability, especially in tasks involving constrained motion near sensitive anatomical structures.

The experimental results provide clear evidence of the proposed framework’s ability to generate safe, accurate, and stable trajectories from a mixed demonstration set. In Cartesian space, the reproduced trajectory remains entirely within the failure-free region, indicating successful avoidance of contact with the wire even in areas with complex curvature. This performance contrasts with that of the GMM and TLFSF baselines, which both exhibit deviation and overshoot, particularly in high-curvature segments. While TLFSF shows partial improvement by suppressing failure-prone regions, it still fails to ensure complete task safety. In terms of orientation, the proposed method achieves smooth and continuous angular transitions across roll, pitch, and yaw. By comparison, both GMM and TLFSF result in irregular and abrupt orientation changes, especially in the roll and yaw axes, which can lead to tool misalignment and operational instability. Together, these demonstrate that the proposed approach not only improves trajectory fidelity but also ensures rotational control, validating its suitability for applications requiring both spatial precision and motion robustness.

The results from both simulation and real-world experiments confirm the effectiveness of the proposed framework for learning from imperfect demonstrations, specifically within the context of the ring-and-rail task. This surgical training task captures key characteristics of surgical subtasks—such as con-

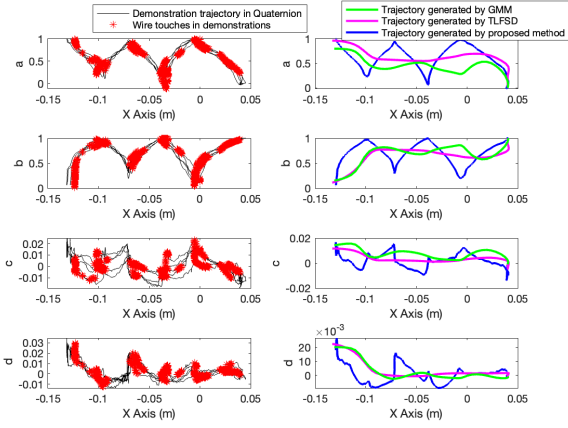


Figure 8: Quaternion representation of end-effector orientation. The proposed method produces a smooth and stable orientation profile throughout the task.

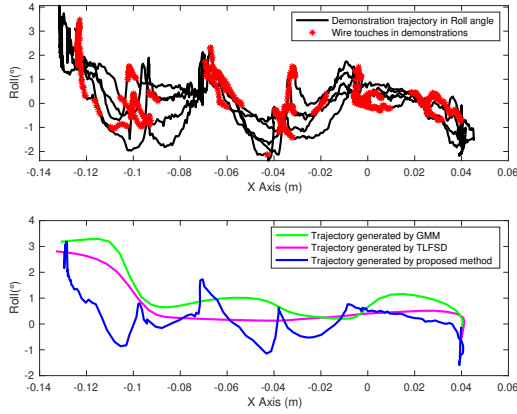


Figure 9: Roll angle of the end-effector. The proposed method maintains stable roll behavior, while baseline methods exhibit discontinuities.

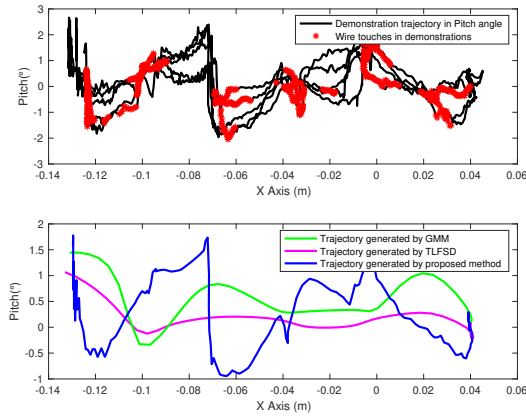


Figure 10: Pitch angle across the trajectory. Smooth transitions reflect consistent control and reduced orientation noise.

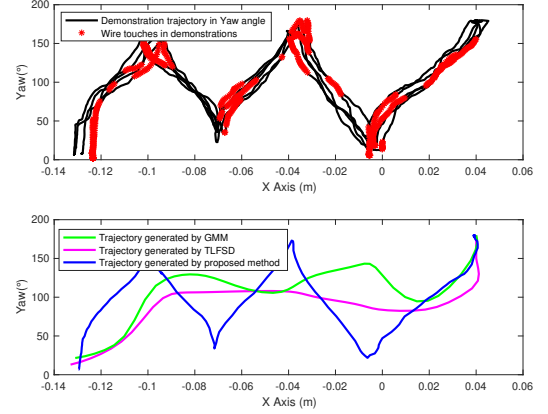


Figure 11: Yaw angle (rotation about the Z-axis). The proposed method avoids abrupt heading changes that could destabilize motion in planar tasks.

strained tool motion, fine motor control, and the need for collision avoidance—and is widely used as a proxy for evaluating robotic surgical skill acquisition. Despite a limited number of demonstrations, including several containing execution errors, the proposed method is able to infer a safe and task-compliant trajectory. The effectiveness of the proposed method suggests its applicability to a wider range of robotic surgical tasks that involve constrained tool motion and heterogeneous demonstration quality. Its robustness to execution variability makes it particularly well suited for both autonomous assistance during surgical procedures and as a training aid for skill acquisition in surgical education.

5. Conclusion and Future Work

This paper presents an optimization-based framework for learning from imperfect demonstrations in robotic surgical tasks. The approach integrates Gaussian Mixture Models (GMMs) to represent successful motion patterns and Donut Mixture Models (DMMs) to encode failure-prone behaviors. These models are incorporated into a multi-objective trajectory optimization formulation that enables the generation of task-compliant and contact-free motions from mixed-quality data. The framework is evaluated through simulation and real-world experiments on the ring-and-rail task, a standard surgical training task for assessing surgical precision and dexterity. Results demonstrate that the method reliably produces safe and smooth trajectories, even when trained on a limited set of demonstrations containing execution errors.

Despite its demonstrated effectiveness, the proposed framework has several limitations. First, it relies on binary failure labels at the time-step level, which currently require task-specific instrumentation such as contact sensors. This dependency may restrict applicability in uninstrumented or more complex surgical environments where failure events are difficult to detect or define reliably. Second, the framework assumes temporally aligned demonstrations, which may not generalize to multi-phase procedures or tasks with variable execution tim-

ing. Third, the current implementation is limited to single-arm, position-controlled tasks with constrained planar motion.

References

- [1] B. T. Ostrander, D. Massillon, L. Meller, Z.-Y. Chiu, M. Yip, R. K. Orosco, The current state of autonomous suturing: a systematic review, *Surgical Endoscopy* 38 (5) (2024) 2383–2397.
- [2] X. Zhu, Y. Jiang, R. He, C. Li, X. Duan, Development and validation of a robot-assisted retraction system for orthopedic surgery, *IEEE Transactions on Medical Robotics and Bionics* (2025).
- [3] H. S. Ghandourah, R. M. Schols, J. A. Wolfs, F. Altaaweel, T. J. van Mulken, Robotic microsurgery in plastic and reconstructive surgery: a literature review, *Surgical innovation* 30 (5) (2023) 607–614.
- [4] K. L. Schwaner, I. Iturrate, J. K. Andersen, P. T. Jensen, T. R. Savarimuthu, Autonomous bi-manual surgical suturing based on skills learned from demonstration, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2021, pp. 4017–4024.
- [5] N. Amirshirzad, B. Sunal, O. Bebek, E. Oztop, Learning medical suturing primitives for autonomous suturing, in: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), IEEE, 2021, pp. 256–261.
- [6] X. Deng, Y. Chen, F. Chen, M. Li, Learning robotic ultrasound scanning skills via human demonstrations and guided explorations, in: 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 2021, pp. 372–378.
- [7] Y. Huang, W. Xiao, C. Wang, H. Liu, R. Huang, Z. Sun, Towards fully autonomous ultrasound scanning robot with imitation learning based on clinical protocols, *IEEE Robotics and Automation Letters* 6 (2) (2021) 3671–3678.
- [8] A. Pore, E. Tagliabue, M. Piccinelli, D. Dall’Alba, A. Casals, P. Fiorini, Learning from demonstrations for autonomous soft-tissue retraction, in: 2021 International Symposium on Medical Robotics (ISMR), IEEE, 2021, pp. 1–7.
- [9] J. W. Kim, C. He, M. Urias, P. Gehlbach, G. D. Hager, I. Iordachita, M. Kobilarov, Autonomously navigating a surgical tool inside the eye by learning from demonstration, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020, pp. 7351–7357.
- [10] I. Rivas-Blanco, C. J. Perez-del Pulgar, C. López-Casado, E. Bauzano, V. F. Muñoz, Transferring know-how for an autonomous camera robotic assistant, *Electronics* 8 (2) (2019) 224.
- [11] O. Brunckhorst, B. Challacombe, H. Abboudi, M. Khan, P. Dasgupta, K. Ahmed, Systematic review of live surgical demonstrations and their effectiveness on training, *Journal of British Surgery* 101 (13) (2014) 1637–1643.
- [12] H. Ravichandar, A. S. Polydoros, S. Chernova, A. Billard, Recent advances in robot learning from demonstration, *Annual review of control, robotics, and autonomous systems* 3 (1) (2020) 297–330.
- [13] D. H. Grollman, A. Billard, Donut as i do: Learning from failed demonstrations, in: 2011 IEEE international conference on robotics and automation, IEEE, 2011, pp. 3804–3809.
- [14] B. Hertel, S. R. Ahmadzadeh, Learning from successful and failed demonstrations via optimization, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2021, pp. 7807–7812.
- [15] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, S. P. DiMaio, An open-source research kit for the da vinci® surgical system, in: 2014 IEEE international conference on robotics and automation (ICRA), IEEE, 2014, pp. 6434–6439.
- [16] P. Zhang, J. W. Kim, P. Gehlbach, I. Iordachita, M. Kobilarov, Autonomous needle navigation in retinal microsurgery: Evaluation in ex vivo porcine eyes, in: 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2023, pp. 4661–4667.
- [17] G. Gonzalez, M. Balakuntala, M. Agarwal, T. Low, B. Knoth, A. W. Kirkpatrick, J. McKee, G. Hager, V. Aggarwal, Y. Xue, et al., Asap: A semi-autonomous precise system for telesurgery during communication delays, *IEEE Transactions on Medical Robotics and Bionics* 5 (1) (2023) 66–78.
- [18] Y. Qin, M. Allan, J. W. Burdick, M. Azizian, Autonomous hierarchical surgical state estimation during robot-assisted surgery through deep neural networks, *IEEE Robotics and Automation Letters* 6 (4) (2021) 6220–6227.
- [19] A. A. Shahkoo, A. A. Abin, Autonomous tissue manipulation via surgical robot using deep reinforcement learning and evolutionary algorithm, *IEEE Transactions on Medical Robotics and Bionics* 5 (1) (2023) 30–41.
- [20] P. M. Scheikl, E. Tagliabue, B. Gyenes, M. Wagner, D. Dall’Alba, P. Fiorini, F. Mathis-Ullrich, Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery, *IEEE Robotics and Automation Letters* 8 (2) (2022) 560–567.
- [21] Y. Ou, A. Soleymani, X. Li, M. Tavakoli, Autonomous blood suction for robot-assisted surgery: A sim-to-real reinforcement learning approach, *IEEE Robotics and Automation Letters* (2024).
- [22] Y. Long, W. Wei, T. Huang, Y. Wang, Q. Dou, Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning, *IEEE Robotics and Automation Letters* 8 (8) (2023) 4441–4448.
- [23] Y. Bi, Z. Jiang, Y. Gao, T. Wendler, A. Karlas, N. Navab, Vesnet-rl: Simulation-based reinforcement learning for real-world us probe navigation, *IEEE Robotics and Automation Letters* 7 (3) (2022) 6638–6645.
- [24] P. M. Scheikl, N. Schreiber, C. Haas, N. Freymuth, G. Neumann, R. Lioutikov, F. Mathis-Ullrich, Movement primitive diffusion: Learning gentle robotic manipulation of deformable objects, *IEEE Robotics and Automation Letters* (2024).
- [25] H. Su, A. Mariani, S. E. Ovrur, A. Mencias, G. Ferrigno, E. De Momi, Toward teaching by demonstration for robot-assisted minimally invasive surgery, *IEEE Transactions on Automation Science and Engineering* 18 (2) (2021) 484–494.
- [26] M.-Z. Pan, Y.-W. Deng, Z. Li, Y. Chen, X.-L. Liao, G.-B. Bian, Dynamic multi-action recognition and expert movement mapping for closed pelvic reduction, *IEEE Transactions on Industrial Informatics* 19 (8) (2022) 8667–8678.
- [27] Z. Cao, D. Sadigh, Learning from imperfect demonstrations from agents with varying dynamics, *IEEE Robotics and Automation Letters* 6 (3) (2021) 5231–5238.
- [28] G. Chou, N. Ozay, D. Berenson, Learning temporal logic formulas from suboptimal demonstrations: theory and experiments, *Autonomous Robots* 46 (1) (2022) 149–174.
- [29] Z. Zhu, K. Lin, B. Dai, J. Zhou, Self-adaptive imitation learning: Learning tasks with delayed rewards from sub-optimal demonstrations, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 36, 2022, pp. 9269–9277.
- [30] D. H. Grollman, A. G. Billard, Robot learning from failed demonstrations, *International Journal of Social Robotics* 4 (2012) 331–342.
- [31] S. Calinon, D. Lee, Learning control, in: *Humanoid robotics: A reference*, Springer Netherlands, 2017, pp. 1–52.