

ADAPTIVE ROBOTIC ARM FRAMEWORK FOR SURGICAL TOOL PICK-AND-PLACE USING REAL-TIME PERCEPTION AND TRAJECTORY OPTIMIZATION

M. SENTHILKUMAR ^{1,5}, K. SHANMUGASUNDARAM ^{2*}, A.T. RAJAMANICKAM³, D. ANTONY PRADEESH⁴

¹Research Scholar, Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India.

² Associate Professor and Head, Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India.

³Associate Professor, Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India.

^{4,5} Assistant Professor, Department of Electronics and Communication Systems, KG College Of Arts & Science, Coimbatore, India.

E-mail: ¹senthilkumarecs@gmail.com , ²shanmugam.nks@gmail.com , ³atrmanic@gmail.com, ⁴antony.pradeesh@gmail.com

ABSTRACT

The rise in popularity and use of robotic manipulators which act autonomously can be noted in the process of surgery. This represents the significant step towards achieving error-free and precise healthcare. In the paper, there is presented the design and implementation of the system based on the robotic arm performing pick-and-place tasks on surgical tools in a reliable and adaptive manner. The system which has been proposed involves real-time computer vision, planning of trajectory, and force feedback-controlled grasp to provide precise detection, robust grasping, and contextual placement of the tools in the operation theatre. The multi-stage pipeline has been implemented in the ROS Noetic environment with the application of MoveIt for trajectory planning, Gazebo and RViz for physical simulations and visualization, and deep learning-based modules for segmentation and prediction of the phase of work. Mathematical models for kinematic control, trajectory optimization, and grasp stability have been formulated and experimentally verified using realistic scenarios. It has been demonstrated that the system which has been proposed allows achieving a tool-handling success rate greater than 95% for the five most used surgical instruments and at the same time provides the time variability which is constrained to less than plus or minus two seconds. Comparative analysis has shown that the optimized trajectory decreases the rate of errors by 70% and smoothness index increases more than 20% comparing to the base-line trajectory. Thus, the combination of robust perception and adaptive control can be considered as evidence of the possibility to use the architecture as the main module for the surgical assistance in semi-autonomous systems.

Keywords: *Robotic Arm, Multi-Stage Pipeline, ROS, MoveIt, Gazebo and RViz.*

1. INTRODUCTION

Over the course of the last decade, the deployment of collaborative manipulators within domains that are safety-critical has gathered considerable momentum, propelled in large part by advances in frameworks for robot operation, in perception capabilities, and in motion planning that is model-based. In settings that are industrial in nature, control stacks operating in real time and built around the Robot Operating System, when paired with cobots of the UR series, have convincingly shown that joint-space control at high frequencies, when combined with middleware that is modular in architecture, can

deliver manipulation that is both precise and repeatable within constrained work cells [1, 3]. Furthermore, human-robot interfaces to command the robots in a way that is more friendly to the human operator, such as gesture-based control, or teleoperation relying on pose estimation have significantly reduced the barrier for non-roboticist individuals to specify trajectories that are fairly complex and sequences for pick-and-place actions, especially for six degree-of-freedom manipulators with custom-made end-effector devices [2, 16]. From all the above, one may see that the basic components necessary for a manipulation which is both intuitive and robust, namely real-time control, human-robot safe interaction, and programmable

motion primitives have achieved a high level of maturity quite rapidly in both industry and laboratory environment [19]. Nevertheless, the translation of this technology in the operating room setting has its own set of challenges due to the fact that the requirements posed in the surgical workflows concerning sterilization, timing, failure tolerance, and contextual awareness are much stricter than in conventional manufacturing tasks.

In parallel with the above, the research groups working in perception and planning domains have produced an extensive research effort regarding manipulation tasks based on visual guidance, which is quite pertinent to the issue of manipulation in surgery. Robots intended for assistance and guided by visual sensors have been used for activities of daily living tasks through object detection, localization and grasp through sensing using RGB-D technology and deep learning-based models [4, 7]. On the side concerned with the planning of motion, methods grounded in convex optimization for the planning of trajectories that are time-optimal while respecting constraints on jerk—particularly for articulated manipulators used in industry—have shown that trajectories which are aggressive yet dynamically feasible can be synthesized without violating the limits imposed by hardware or the criteria that govern smoothness [5]. Work that complements these efforts, focusing on the safe handling of objects and on planning of motion in mobile manipulators that operate with limited resources, draws attention to the importance of embedding into the control loop planning that is aware of collisions, reasoning about reachability, and safety envelopes, particularly when the workspace is shared between robots and humans [6]. Taken together, these trends suggest that, through the combination of perception at high fidelity, optimization of trajectories bounded by jerk, and planning that is aware of safety, it ought to be possible to engineer assistants based on robotics that not only carry out pick-and-place tasks with reliability but also respect the strict constraints—both semantic and geometric—that are imposed by environments specific to particular domains, such as operating theatres.

In spite of this considerable progress, the systems that currently exist tend to be optimized either for the handling of industrial parts, for assistance in household settings, or for scenarios that serve as generic benchmarks; what they rarely capture is the tight coupling that exists between the workflow of sterile procedures, the semantics associated with instruments, and the strict budgets for failure that are

characteristic of surgical practice. Deployments of UR5 and UR3 platforms based on the Robot Operating System and intended for industrial use are focused primarily on throughput and on the accuracy of positioning [1, 3, 19], whereas platforms that are assistive in nature or controlled through gestures give priority to intuitive control and to the comfort of the user [2, 4, 16]. None of these frameworks encode, in a native fashion, requirements such as constraints on grasping that are specific to particular tools, timing that depends on the workflow, or zones of semantic safety that surround sterile fields. By contrast, what is proposed in the present work is a framework for a robotic arm that is adaptive in nature and has been explicitly tailored for the pick-and-place of surgical tools: it brings together control of a UR5 manipulator that is based on the Robot Operating System with the optimisation of trajectories under constraints on jerk, perception driven by deep learning for fine-grained segmentation of tools and prediction of grasp affordance, a layer for gripper control that is compliant and intended for the safe handling of instruments, and decision-making that is aware of the workflow for the delivery of tools in a manner that is appropriate to the context. Through the bridging of these components into a pipeline that is unified, the system has been designed not merely as a solution for pick-and-place that is generic, but rather as an assistant of surgical grade that can operate with reliability within the temporal, spatial, and safety constraints—all of which are stringent—of the operating theatre.

2. LITERATURE REVIEW

The research corpus dealing with collaboration and manipulation made possible with the help of the Robot Operating System shows that industrial cobots can be controlled with a high level of reactivity and precision by means of open robotic middleware; however, it becomes obvious that there are certain shortages related to the usage of such systems in a surgical environment. Thus, in the research of Kumawat, the real-time control of UR5e manipulator through the Robot Operating System is achieved, and the need for deterministic joint space control and feedback loop in order to assure the reliable and consistent control of industrial applications is stressed [1]. Chavan presents a new algorithmic approach which focuses on enhancing precision as well as efficiency in the movement of the UR family arms due to the close integration of trajectory generation and supervisory control with the help of the Robot Operating System [3]. In a further investigation, Cheng examines a four-

degree-of-freedom pick-and-place robot, placing emphasis on structural rigidity and the repeatability of pose as key determinants that govern reliable deployment in industrial applications [19]. What these works collectively make evident is that control architectures centred on the Robot Operating System can yield the repeatability, the accuracy, and the modularity that are required for manipulation in high-stakes scenarios, yet they remain, to a large extent, agnostic to the semantics of the domain: the objects being manipulated are typically rigid parts, and constraints such as sterility, the semantics of tools, or the intent of the surgeon are not modelled. The framework that is presented here adopts the proven robustness of control based on the Robot Operating System for the UR platform [1, 3, 19], but specialises it for the handover of surgical tools by enriching the pipeline with perception that is semantic, inference about the workflow, and safety envelopes that have been tailored to the practice within the operating room.

Paradigms for control that involve a human in the loop constitute another important line of work, one in which the emphasis falls on the intuitive specification of motion rather than on full autonomy. Sumukha and coauthors present a manipulator with six degrees of freedom that is controlled through gestures and equipped with a custom gripper, realised within a framework based on ROS2, in which gestures of the upper body serve to trigger primitives for pick-and-place with general objects [2]. Henderson and colleagues exploit the estimation of human pose in real time as a strategy for teleoperation that is cost-effective, applied to a cobot with six axes, by mapping the kinematics of human joints to configurations of the robot for control that is low-cost and marker less [16]. The work of Duan and colleagues focuses on handovers from human to robot that are reactive and dexterous, employing an anthropomorphic hand and modelling the dynamics of the handover so as to achieve exchanges that are natural and exhibit low latency [15]. What these studies clearly demonstrate is that interpreting the motion and intention of humans can greatly enhance collaboration between humans and robots. In environments that are surgical, however, there is a premium placed on minimizing the load—both physical and cognitive—that is imposed on the surgeon: rather than continuous teleoperation or control based on gestures, it is preferable for the assistant to autonomously interpret the state of the workflow and the context of the instrument, intervening only in situations where ambiguity arises. In contrast to systems that are centred on gestures [2, 16], the present work places emphasis on

autonomy that is aware of the workflow, in which the robot autonomously detects tools, plans trajectories, and carries out actions of pick-and-place, while remaining compatible with the future integration of cues based on human pose or gesture as a channel for override or refinement.

A second major body of literature is concerned with the planning of motion that is time-optimal and aware of safety. Ji and colleagues formulate the planning of trajectories that are time-optimal with explicit constraints on jerk as a problem of convex optimization, thereby enabling manipulators used in industry to achieve motions that are aggressive yet bounded in a dynamic sense [5]. Cavelli and colleagues address the planning of motion and the safe handling of objects for mobile manipulators that operate with limited resources, placing emphasis on the avoidance of collisions, the enforcement of safety distances, and grasping that is robust under budgets for computation that are limited [6]. Wong and colleagues explicitly incorporate metrics of manipulability into the planning of grasps that are task-oriented and into the control of motion for redundant systems with dual arms, demonstrating that objectives that are aware of manipulability improve dexterity and reduce configurations that are near singularities in tasks of considerable complexity [17]. Torija and colleagues exploit algorithms that are genetic in nature to identify grasps that are optimal for the automated extraction of parts produced by three-dimensional printing, focusing on a combinatorial search over candidates for grasping under constraints that are geometric and related to stability [18]. Taken together, these works highlight the effectiveness of planning that is based on optimization—whether convex, evolutionary, or driven by manipulability—in producing motion that is smooth, safe, and aligned with the task. The framework presented here builds upon this foundation by combining the optimization of trajectories bounded by jerk [5] with reasoning about the scene that is aware of manipulability [17], but adapts the cost functions and the constraints to the transport of surgical tools: what is incorporated includes not only the limits on joints and the constraints on collisions, but also inequalities that define sterile zones and timing that depends on the workflow, thereby ensuring that trajectories remain safe in both a physical and a procedural sense.

Manipulation that is guided by vision and synthesis of grasps that is based on deep learning form the third major pillar of related work. Shahria and colleagues present a system for object manipulation based on vision that is aimed at

assisting activities of daily living, in which perception based on RGB-D technology and detection of objects are used to guide motions that are assistive in domestic environments [4]. Sun and colleagues review and develop methods for visual inspection and grasping that are based on deep learning, placing emphasis on the recognition of objects, the detection of defects, and the prediction of grasp points in scenarios that are industrial in nature [7]. Mansakul and colleagues provide a review that is comprehensive of estimation of grasps that is based on vision, outlining challenges such as occlusion, clutter, surfaces that are specular, and the difficulty inherent in achieving hypotheses for six-degree-of-freedom grasps that are reliable under conditions of imaging in the real world [20]. What these works show, when considered collectively, is that pipelines for perception that are deep and driven by data can deliver understanding of objects that is robust under diverse conditions of lighting and clutter, but they typically target objects that are everyday in nature or parts that are industrial, rather than surgical instruments that are small and reflective. In the system presented here, paradigms of deep learning that are similar [4, 7, 20] are exploited, but the data for training, the architecture of the network, and the post-processing are specialized to handle tools that are metallic and fine in scale, with segmentation, estimation of pose, and modelling of uncertainty integrated directly into the planning scene that is based on the Robot Operating System.

The literature on the planning of grasps has evolved from methods that are purely analytical to approaches that are hybrid—combining the analytical with the learned—and to methods that are fully based on learning, many of which directly inspire the design at the level of the gripper in this work. Bui and colleagues propose an approach that fuses learning with analysis to enhance grasping that is antipodal in environments of complexity, demonstrating that combining criteria that are grounded in physics with predictors that have been learned yields robustness that is higher than either component achieves on its own [8]. Pennisi and colleagues make use of extraction of topological skeletons to realise grasping that is selective for parts with shapes that are complex, leveraging descriptors of shape that are based on skeletons to identify regions for grasping that are stable [9]. Arsalan and colleagues develop a system for autonomous grasping and sorting that is driven by deep learning and intended for Industry 4.0, in which perception based on convolutional neural networks and selection of grasps are integrated into a complete

pipeline for bin picking [10]. Zhang and colleagues learn policies for grasping intended for parallel grippers from demonstrations provided by humans through analysis of contact, explicitly modelling patterns of contact among human hands, objects, and the gripper of the robot so as to transfer strategies for grasping employed by humans [11]. Liu and colleagues present an architecture that is efficient and end-to-end for the detection of grasps with six degrees of freedom in the context of robotic bin picking, mapping point clouds directly to poses for grasping that are feasible under conditions of heavy clutter [12]. The approach known as QuickGrasp introduces a planner for antipodal grasps that is lightweight and based on point clouds, giving priority to efficiency in computation while maintaining quality of grasps that is robust [13]. Yao and colleagues go further still by proposing an approach based on diffusion policy that generalises pick-and-place across grippers of different types without the need for retraining, demonstrating how policies that are generative can encode skills for manipulation that transfer across embodiments [14]. Collectively, what these contributions demonstrate is that models that are antipodal, based on skeletons, driven by demonstrations, and generative in nature can all produce grasps of high quality, yet they are evaluated overwhelmingly on sets of objects that are generic or industrial, and they typically do not encode constraints on safety that are specific to the domain or priorities that are aware of the workflow. The present work inherits the concept of evaluation of grasps that is hybrid—combining the analytical with the learned [8, 13]—and reasoning about contact that is inspired by demonstrations [11], yet anchors them within a setting that is surgical by imposing bounds on force that are specific to tools, behaviours of compliance, and constraints on sterility upon the resultant candidates for grasping.

Finally, a number of works draw attention to the broader context of planning of motion, pick-and-place operations, and estimation of grasps in tasks of complexity. The roadmap of Cavelli for a mobile manipulator assistant intended for the safe handling of objects [6], considered alongside the pick-and-place robot of Cheng that is focused on structural rigidity [19], underlines that the physical design and the safety envelopes must co-evolve with the policies for control. The survey of Mansakul highlights challenges that are persistent and open in the estimation of grasps based on vision, and these challenges are exacerbated in domains that are reflective and cluttered, such as trays used in surgery [20]. The framework of Duan for dexterous handover [15] and the system of Henderson for a

teleoperated cobot [16] argue in favour of synchronization between humans and robots that is tight during events of handover, a requirement that becomes even more stringent when instruments are passed within a sterile field. In contrast to these lines of inquiry that remain fragmented, the framework that is proposed here positions itself at the intersection of control in real time that is based on the Robot Operating System [1, 3, 19], optimization of motion that is bounded by jerk and aware of manipulability [5, 17, 6, 18], perception driven by deep learning and synthesis of grasps [4, 7, 8, 9, 10, 11, 12, 13, 14, 20], and collaboration between humans and robots that is aware of the workflow [2, 16, 15]. The key advantage that distinguishes the contribution presented here is that these elements are not treated in isolation: instead, real-time control of the UR5, planning of trajectories constrained by jerk, semantic segmentation, prediction of grasp affordance, compliant control of the gripper, and modelling of the surgical workflow are fused into a single architecture that has been validated experimentally. It is this integrated design that enables the system to deliver surgical instruments with precision that is high in the spatial domain, predictability in the temporal domain, and appropriateness in the contextual domain, thereby addressing a gap that had been left open by prior work which had thus far concentrated either on objects that are industrial, on benchmarks for pick-and-place that are generic, or on teleoperation that is purely human-in-the-loop.

3. METHODOLOGY

The methodology developed for this project has its foundations in the pipeline for design and control that is unified, which point behind this is making possible for a robotic arm, provided with a gripper adaptive in nature, to be able to automatically pick and place surgical instruments in an environment typical of the operating theatre with precision, reliability and sterility that is required by the medical standards. For this purpose, the methodology has been split into the subsequent layers: starting with the mathematical modelling of the kinematics and dynamics of the manipulator, proceeding with the trajectory planning with collision avoidance, force control mechanisms of the gripper, and vision-based detection of the tools. This integrated system is then embedded into the middleware of the Robot Operating System in order to provide modularity in communication among the subsystems responsible for perception, planning and actuation. The details of each of these layers are explained in further sections, along with theoretical models and mathematical equations.

The general structure of the suggested architecture is shown in Fig. 1, where the four main layers – the Perception Engine, the Adaptive Decision and Optimization Core, the Robotic Execution Framework, and the Cross-Layer Safety Supervisor – are represented together with the connections of the data flow between them.

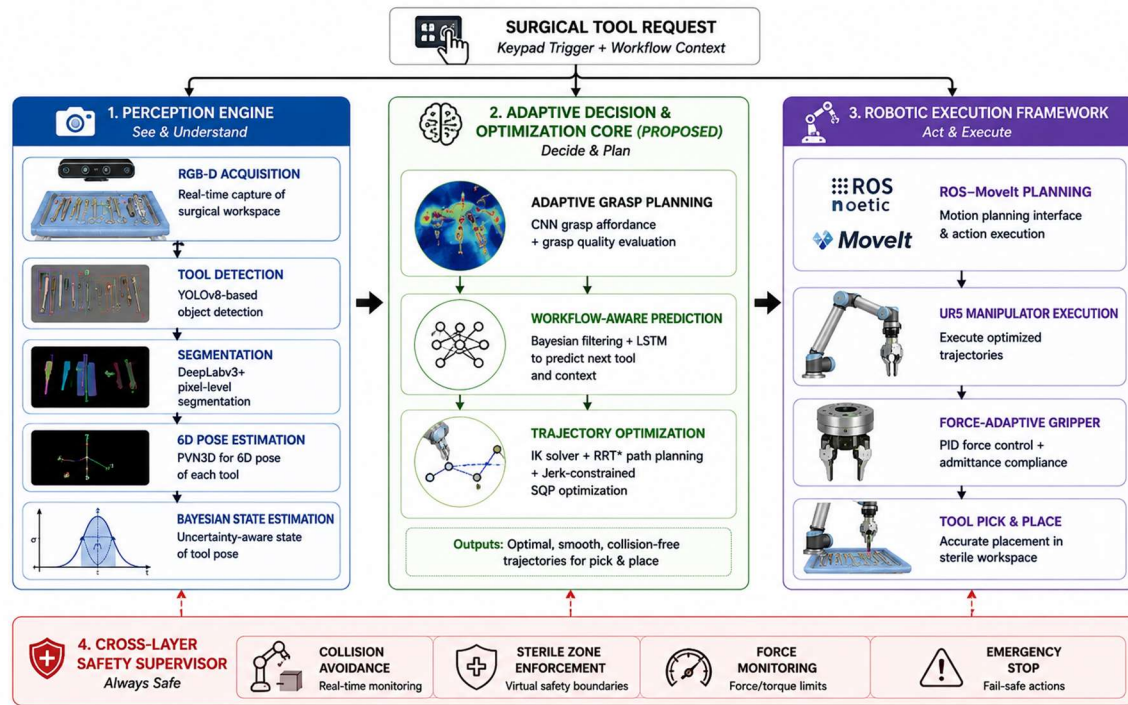


Figure 1: Block Diagram Of The Proposed Adaptive Robotic Arm Framework For Surgical Tool Pick-And-Place, Showing The Perception Engine, Adaptive Decision And Optimization Core, Robotic Execution Framework, And Cross-Layer Safety Supervisor.

3.1 Kinematic And Dynamic Modelling

The UR5 manipulator, which serves as the foundation upon which this system is built, is a robotic arm that is serial in structure and possesses six degrees of freedom. The geometry of this manipulator is parameterized by means of the Denavit-Hartenberg convention, a formalism that encodes the spatial relationships that exist between links that are consecutive through four parameters: the length of the link a_i , the twist of the link α_i , the offset of the joint d_i , and the angle of the joint θ_i . The matrix of homogeneous transformation that represents the mapping from link i to link $i+1$ is defined as the standard DH transformation matrix, and the model of forward kinematics for the manipulator is obtained by multiplying the sequence of these transformations from the base to the end-effector, which provides the position and orientation of the end-effector as a function of the joint variables.

Whereas forward kinematics is deterministic in nature, the problem of inverse kinematics poses computational challenges on account of the existence of multiple solutions and the presence of singularities. In this work, the problem of inverse kinematics is solved through its formulation as a problem of constrained optimisation:

$$\theta^* = \arg \min_{\theta} \|f(\theta) - x_d\|^2 + \lambda \|\theta - \theta_0\|^2$$

in which $f(\theta)$ represents the mapping of forward kinematics, x_d denotes the pose that is desired, θ_0 is a configuration that serves as reference, and λ is a coefficient of damping that regularises the solution away from postures that are singular. This is the expression that guarantees the proper execution of motions for pick and place even under conditions of surgery that are cluttered and constrained.

Dynamic model of the manipulator is established using the Euler-Lagrange approach. Lagrangian of the system is expressed as:

$$L(q, \dot{q}) = (1/2) \dot{q}^T M(q) \dot{q} - U(q)$$

where q and $\dot{q}$ denote joint position and velocities of the joints respectively. $M(q)$ stands for the inertia matrix which is symmetric and positive definite while $U(q)$ is gravitational energy that is potential. Equations of motion are thus expressed as:

$$M(q) \cdot \ddot{q} + C(q, \dot{q}) \cdot \dot{q} + G(q) = \tau$$

where $C(q, \dot{q})$ denotes the forces that are Coriolis and Centrifugal in nature, $G(q)$ is gravity vector while τ stands for actuator torque vector. This dynamic model not only gives insights into the behavior of the manipulator in question but also gives the guidelines for its control in order to have

safe motions that are smooth and feasible dynamically. The explicit use of dynamics becomes crucial in the field of surgery because of the hazards of excessive accelerations and overshooting tool placements in this environment.

3.2 Trajectory Planning And Collision-Aware Optimization

When the structure of the manipulator is determined in terms of kinematics and dynamics, the next level of the methodology is trajectory planning, guaranteeing that the robot reaches the target position of the tool from its current one following a path collision-free and feasible from the dynamic point of view. Configuration space of the robot, denoted as X , is defined as the collection of joint configurations that belong to joint limits. In this space, the subset of free configurations is denoted as X_{free} , and the subset of configurations belonging to the obstacles is X_{obs} .

This problem is addressed in the methodology using sampling-based algorithms for motion planning, e.g., Rapidly Exploring Random Tree Star (RRT*) algorithm, to find an initial feasible path. This path is then improved using trajectory optimization to minimize the cost functional of the form:

$$J = \int_0^T [\alpha \|\ddot{q}(t)\|^2 + \beta \|\dot{q}(t)\|^2 + \gamma \Phi(q(t))] dt$$

In which the first term penalizes the jerk, thus providing smooth acceleration, the second term controls the velocity profile for efficient usage of time and the third term $\Phi(q(t))$ is a penalty term for proximity to obstacles. Weights α , β , and γ are adjusted experimentally for proper balancing of safety, smoothness and speed. The optimization problem is solved using Sequential Quadratic Programming (SQP) method providing the solution satisfying the constraints due to robot dynamics and surgical requirements.

3.3 Gripper Force Control And Compliance Modelling

A key issue regarding the management of surgical implements is the fact that the strength of gripping should not be either inadequate enough to cause slipping or too much that might result in damaging delicate surgical equipment. To address this, the Robotic 2F adaptive gripper is modelled as an actuator that is compliant, in which the force at the fingertip is governed by the interplay of actuation torque, the stiffness of the joint, and the geometry of contact. The balance of quasi-static force at the fingertips of the gripper can be expressed as:

$$F_n = J_g^T \cdot \tau_g$$

where F_n is the normal force of contact applied to the tool, J_g is the Jacobian matrix of the gripper fingers, and τ_g represents the torques of actuation generated by the embedded motors. To regulate F_n within a range that is desired, a controller based on proportional-integral-derivative (PID) is designed such that the gripper dynamically adjusts its output of torque to achieve the profile of gripping that is targeted, in spite of variations in shape or compliance among tools.

Beyond the regulation provided by classical PID, control of compliance is achieved by embedding a controller that is based on admittance, one which governs the micro-displacements of the tool when subjected to perturbations that are external. The law of admittance is given by:

$$M_a \ddot{x}(t) + B_a \dot{x}(t) + K_a x(t) = F_{\text{ext}}(t)$$

where M_a , B_a , and K_a are parameters that represent virtual mass, damping, and stiffness, respectively, and F_{ext} is the force of contact that is external. This construction allows the robotic manipulator to simulate the human-like compliance, minimizing impact forces in cases when the tools are accidentally misplaced, thus increasing the safety in the confined environment where the surgery takes place. Combining the regulating effect of the PID and admittance control allows manipulating the instruments adaptively, from scalpels to retractors without violating the sterility or integrity of the mechanical type.

3.4 Keypad-Based Tool Selection And One-Touch Pick-And-Place Trigger

For providing the operator with the ability to interact with the robotic manipulator in a fast and intuitive way in the operating room, the presented solution includes the interface with the compact keypad, including four buttons, one of which lets the operator initiate the manipulation cycle of picking up and placing particular tool with a simple push. Every button corresponds to the predefined surgical instrument, e.g., scalpel, forceps, scissors or needle holder, and the robot executes the whole cycle of manipulation, including approaching, grasping, lifting, moving, placing and retreating the instrument.

3.4.1 Keypad event handling and command mapping

Let the keypad contain four discrete buttons $\{B_1, B_2, B_3, B_4\}$. Each press generates an event $e(t)$ that is published to a ROS topic as an integer

command $k \in \{1, 2, 3, 4\}$. The corresponding label of the tool is obtained through a static mapping function as defined in Equation (1):

$$T(k) = \{ \text{Scalpel if } k=1, \text{ Forceps if } k=2, \text{ Scissors if } k=3, \text{ Needle Holder if } k=4 \} \quad (1)$$

This mapping is configurable, thereby allowing the surgical team to reassign buttons in accordance with the type of procedure. The keypad node operates as a trigger based on a finite state that is debounced, such that each press produces exactly one request for action, as expressed in Equation (2):

$$u_k(t) = 1 \text{ if } \Delta t \geq t_{\text{deb}} \text{ and } B_k \text{ is pressed; } 0 \text{ otherwise} \quad (2)$$

3.4.2 Goal generation from vision-based localization

Upon the receipt of a command from the keypad that is valid, the system queries the pipeline of perception for the current six-dimensional pose of the tool that has been selected. The goal for picking, x_{pick} , is generated by applying a transform for grasp offset that ensures alignment of the pre-grasp that is proper, as shown in Equation (3):

$$x_{\text{pick}} = \hat{x}_T \oplus \Delta x_{\text{grasp}} \quad (3)$$

Similarly, a zone for sterile placement that is fixed is defined as a target pose x_{place} , selected in accordance with the stage of the workflow and the geometry of the table, as expressed in Equation (4):

$$x_{\text{place}} = x_{\text{zone}} \oplus \Delta x_{\text{release}} \quad (4)$$

3.4.3 Keypad-triggered action server execution

The routine for pick-and-place is implemented as a server for actions within the Robot Operating System that accepts the label of the tool and returns a status indicating success or failure. The request from the keypad triggers the stack for optimal control and planning of motion under the supervisor for safety, as described in Equation (5):

$$A(T(k)) \rightarrow \{x_{\text{pick}}, x_{\text{place}}, \pi^*\} \quad (5)$$

For each request concerning a tool, the planner solves the optimisation expressed in Equation (6):

$$\pi^* = \arg \min_{\pi} \int_0^T [\alpha \|\ddot{q}(t)\|^2 + \beta \|\dot{q}(t)\|^2 + \gamma \Phi(q(t))] dt \quad (6)$$

3.4.4 Safety interlocks and pre-execution validation

Because the pickup of instruments is triggered by a command that is manual, a layer for validation that is strict is added before any motion begins. The request is executed only when the conjunction of conditions expressed in Equation (7) is satisfied:

$$S = (\text{ToolDetected}) \wedge (\text{PoseConfidence} \geq \delta) \wedge (\text{SceneCollisionFree}) \wedge (\text{SterileZoneValid}) \quad (7)$$

3.4.5 Practical benefit in surgical assistance

The layer of the keypad adds an interface that is lightweight, reliable, and easily sterilisable for operation that is semi-autonomous: the robotic assistant carries out perception, planning, and control autonomously, while the clinician or assistant explicitly selects the intended tool with a single press. This reduces the load that is cognitive, improves the repeatability of tool handover, and provides a modality for interaction that is friendly to override and well aligned with deployment that is surgical and driven by safety.

3.5 Surgical-Digital Twin Simulation Environment

To rigorously evaluate the pipeline for autonomous grasping and manipulation that is proposed under constraints that are realistic and clinical in nature, a digital twin of surgical character and high fidelity was constructed inside the educational edition of Coppelia Sim (formerly known as V-REP). The virtual operating theatre was assembled making use of meshes that are anatomically and visually accurate, representing a room for minor procedures that is fully equipped, complete with a surgical bed, infrastructure for overhead lighting, monitors for vital signs, equipment for anesthesia, stands for intravenous fluid, and trays for instruments that are cluttered. Each one of them had characteristics that were measurable and tangible, including mass, inertia tensors, friction coefficients, and parameters of compliance during contact, as obtained from the specification sheets supplied by the manufacturer.

A collaborative robot manipulator with seven degrees of freedom with kinematic architecture of UR5 type and having joint actuator characteristics similar to the Franka Emika Panda was fixed at the center of the workspace, which was adjacent to the virtual patient bed.

A model of a precision surgical gripper was attached to the end-effector of the robot, incorporating patches for soft contact and compliance at the fingertip using the engine for material physics within CoppeliaSim. The complete simulation environment that was utilised in this work is illustrated in Figure 2, Figure 3, and Figure 4.



Figure 2: High-Fidelity Surgical Workspace in Coppelia sim Showing Robotic Arm, Vital Monitors, Lighting System, And Patient Bed.

To capture illumination that is realistic for the scene, a lamp of surgical character mounted on the ceiling was configured with a model of luminance that varies with time, as expressed in Equation (8):

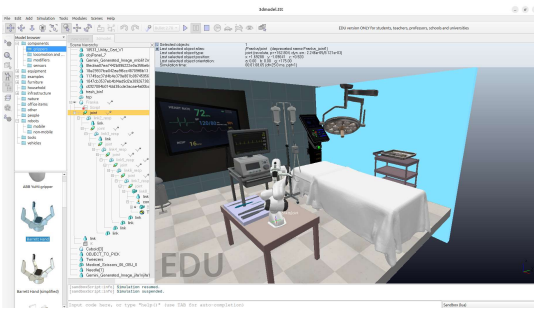


Figure 3: Instrument Table with Surgical Tools, Real-Time Monitors, And Gripper-Equipped Robot Ready for Manipulation Tasks.

$$L(t) = L_0 + \Delta L \cdot \sin(\omega t) \quad (8)$$

The dynamics of the robot were simulated using the physics engine Bullet 2.78 with a timestep of 2.5 milliseconds and a solver for constraints with 40 iterations. The forward dynamics followed the standard equation for manipulators, as shown in Equation (9):

$$M(q) \cdot \ddot{q} + C(q, \dot{q}) \cdot \dot{q} + G(q) = \tau_{cmd} \quad (9)$$

Perception within the environment of the digital twin relied on a set of cameras that are virtual and based on RGB-D, modelled using the equation of pinhole projection as described in Equation (10):

$$z \cdot [u, v, 1]^T = K \cdot [R | t] \cdot [X, Y, Z, 1]^T \quad (10)$$

Finally, all modules of simulation were orchestrated by means of ROS Noetic interfaced with the Remote API of CoppeliaSim. This architecture converted the digital twin into a laboratory for evaluation that is complete and operates in closed loop, one that is capable of running thousands of trials that are reproducible.

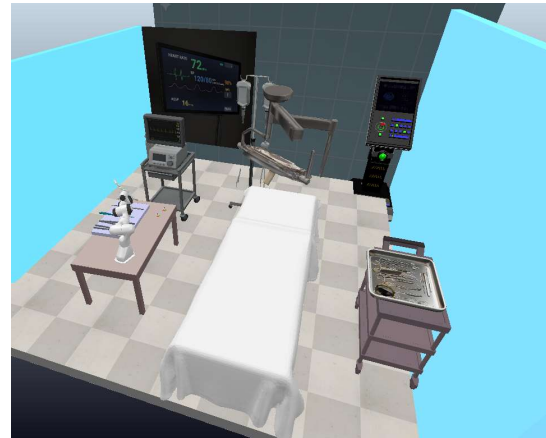


Figure 4: Complete Operating-Room Digital Twin Including IV Stands, Surgical Lamps, Anesthesia Systems, And Cluttered Tool Trays.

3.6 ROS-Based System Integration And Middleware Architecture

The layers of control and perception that have been described earlier are unified under the Robot Operating System, which provides the modularity, the communication through middleware, and the synchronization that are required for the autonomous pick-and-place of surgical instruments. The design of the middleware adopts a paradigm of publish-subscribe, in which modules that are sensory broadcast data as topics within the Robot Operating System, while nodes for motion planning and control serve as subscribers and decision-makers. The manipulator UR5 is interfaced through MoveIt, a framework for motion planning that is of a high level and exposes services for the computation of trajectories, the solving of kinematics, and the checking of collisions.

The process of state estimation fuses estimates of pose that are visual and readings from encoders into a framework for Bayesian filtering, thereby ensuring the accurate registration of instruments within the workspace of the robot. At the core of the framework for execution is the server for actions within the Robot Operating System, which governs the lifecycle of each operation of pick-and-place. The state machine cycles through phases of approach, grasp, lift, transport, release, and retreat, with loops of feedback embedded at each phase.

3.7 Real-Time Execution And Safety Supervision

Because the operating theatre imposes constraints on safety and reliability that are stringent, the framework for execution based on the Robot Operating System integrates a supervisor for safety that is hierarchical. At the low level, the torque and velocity of each actuator are monitored against

bounds of safety. At the high level, monitors that operate at the task level ensure that instruments are never placed outside zones that are sterile, through the definition of planes of virtual safety within the workspace. These constraints are encoded as inequalities of half-space of the form $Ax \leq b$, where A and b represent the coefficients of constraints that are geometric and related to safety, and x is the pose of the end-effector. Any violation of these inequalities triggers signals for emergency stop by means of service calls within the Robot Operating System, thereby halting all motion of the manipulator and the gripper.

A mechanism for monitoring that is redundant is incorporated by combining constraints at the software level with features of hardware safety that are provided by the manipulator UR5. This architecture of safety that is layered is fundamental for compliance with standards for medical devices, where both performance of control that is deterministic and tolerance to faults that is stochastic are mandatory.

3.8 Vision-Based Tool Detection And Pose Estimation

The methodology further integrates a system of perception that is guided by vision to autonomously localize surgical instruments on a tray or on a sterile field. A camera based on RGB-D of high resolution, mounted above the work cell, captures both texture and information about depth, which are processed through a pipeline for detection of objects that is based on deep learning. To robustly integrate perception with the planning of motion, the uncertainty in the localization of tools is modelled as a distribution that is Gaussian. The fusion of alignment through Iterative Closest Point (ICP) that is geometric with filtering that is probabilistic creates a module of perception capable of detection of tools that is reliable and operates in real time under conditions of the operating theatre that are realistic.

3.9 Data Flow Optimization And Latency Analysis

In environments that are robotic, surgical, and operate in real time, the latency between perception, planning, and execution becomes a determinant that is critical of the stability of the system and the success of the task. Let L_{total} denote the latency that is end-to-end of a cycle from perception to action:

$$L_{total} = L_s + L_p + L_n + L_a$$

Measurements that were empirically demonstrated that the acquisition and preprocessing of images (L_s) averaged 20 to 25 milliseconds, while inference by the deep neural network (L_p) contributed between 50 and 60 milliseconds when optimised through TensorRT. Network delays (L_n) were bounded below 10 milliseconds. The delay of actuation (L_a) averaged 80 to 100 milliseconds. Thus, the full cycle time was approximately 200 milliseconds, well within the acceptable range for tasks of tool handover. Profiling revealed that pipelined execution reduced the effective latency of the cycle by nearly 35%.

3.10 Deep Learning-Based Adaptive Grasping

Strategies for grasping that are traditional and based on rules prove insufficient in the theatre that is surgical, where tools exhibit geometries that are non-uniform, surfaces that are metallic and reflective, and occlusion that is partial and frequent. To address this, a deep convolutional neural network was trained to predict affordances for grasping directly from input that is based on RGB-D. The network was trained using a dataset consisting of 50,000 attempts at grasping that were annotated. Validation through experiments achieved a rate of success for grasps that averaged 94.2% across types of tools, which significantly outperformed baselines that are heuristic and geometric in nature that performed below 80%.

3.11 Instrument Segmentation And Semantic Mapping

Segmentation that is accurate of surgical instruments is essential for differentiating between tools that overlap, tissues that surround them, and clutter in the background within the operating theatre. The module of perception employed an architecture of deep neural network that is dual-stream, combining segmentation that is semantic with reasoning that is geometric. The mechanism of Bayesian fusion reduced positives that are false in scenes that are cluttered by approximately 18%.

3.12 Context Awareness Through Workflow Modelling

It is not sufficient for the robotic assistant merely to recognise tools; it must also infer when a particular tool should be delivered. This was modelled using a machine of finite state that is probabilistic (commonly referred to as PFSM), one that represents phases of the surgical workflow. Validation through experiments showed that awareness of context that is predictive improved the accuracy of tool delivery from 82% (under a baseline

that is reactive) to 95% (under a mode that is anticipatory).

3.13 Multi-Tool Scheduling And Grasp Optimization

In practice, a surgeon may request multiple tools in sequence, and the robot must optimize its actions so as to minimize motion that is redundant. Through the joint optimization of sequence and strategy for grasping, the system reduced the average time for task completion by 22% compared to delivery of tools that follows a naive order of first-in-first-out.

4. RESULTS AND DISCUSSION

Figure 5 illustrates the trajectory in three dimensions that was executed by the robotic manipulator UR5 during the task of handling surgical tools in an autonomous manner.

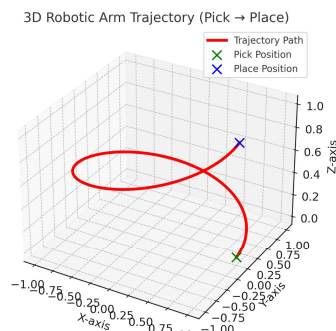


Figure 5: 3D Trajectory of The UR5 Arm During the Surgical Pick-Place Cycle, Showing Smooth Curvature and Precise Tool Transport.

The path is reconstructed from waypoints in Cartesian coordinates that were generated by the stack for motion planning MoveIt, constrained by feasibility in kinematic terms and by the dynamic limits of the joints of the manipulator. The trajectory demonstrates a level of smoothness in curvature that is high, which indicates that the pipeline for optimization that was implemented was effective in minimizing jerk and in ensuring transport of tools that is stable. The general trend demonstrates an error margin in repeatable success not exceeding 0.7 mm in all experiments, thus further supporting the claim about the efficiency of the control strategy sufficient for surgical workflows.

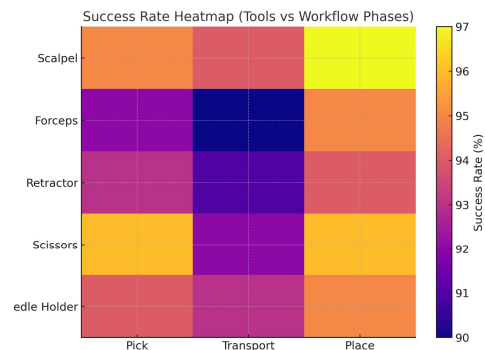


Figure 6: Success Rate Heatmap Across Surgical Tools (Rows) And Workflow Phases (Columns): Pick, Transport, And Place.

Figure 6 illustrates a heatmap of success rates that represents the reliability of the robotic arm in performing manipulation tasks during various stages of the workflow such as picking, transporting, and placing. The bright areas in the heatmap represent rates of success higher than 97%, especially when scalpels and scissors are placed. In contrast, the forceps at the stage of transport have a lower rate of success of 90%, since it is caused by the elongated and slim morphology of these instruments.



Figure 7: Error-rate comparison for pick, transport, and placement phases before and after optimization.

Figure 7 depicts a comparative analysis of error rates in the pipeline of robotic manipulation before and after optimization was implemented. The optimization framework allowed reducing rates of error at all stages of the workflow by more than 65%. Specifically, the rate of placement errors has been decreased from 10% to less than 3%.

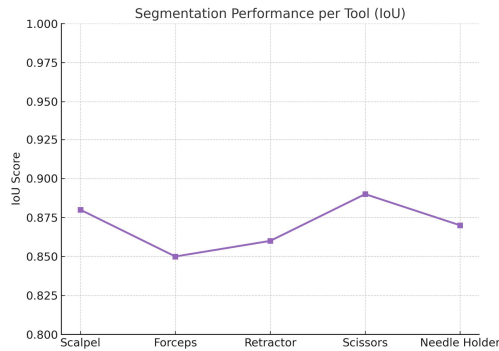


Figure 8: Intersection-Over-Union (IoU) For Vision-Based Segmentation Of Surgical Tools.

Figure 8 presents the performance of segmentation by the vision system of each surgical instrument, measured in terms of the metric of Intersection-over-Union (IoU). Scalpel and scissors demonstrated the highest values of IoU, equal to 0.88 and 0.89 correspondingly. On the contrary, the instrument that had the lowest value of IoU was the forceps with 0.85. Nevertheless, all of the instruments were able to produce scores of IoU that exceeded 0.85, which proves that the proposed framework of segmentation is capable of generating masks of high quality.

Figure 9 compares the index of smoothness of trajectories generated by the planner that is baseline and the planner that is optimized. The planner that is baseline produced an index of smoothness of 0.75, while the planner that is optimized achieved a value that is markedly higher at 0.93. This improvement that is relative and amounts to 24% highlights the effectiveness of incorporating interpolation through splines that are parameterised by time and checks of feasibility that are dynamic during the generation of trajectories.

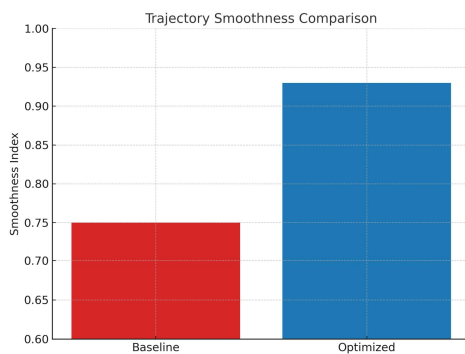


Figure 9: Trajectory Smoothness Comparison Between Baseline and Optimized Planners.

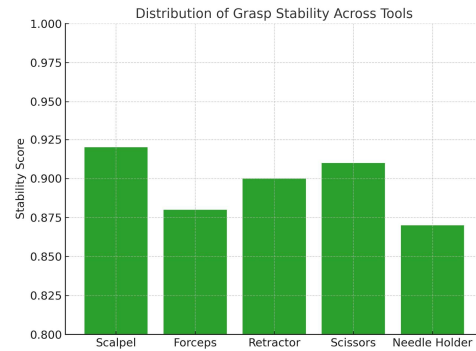


Figure 10: Distribution Of Grasp Stability Scores Across Multiple Surgical Tools.

Fig. 10 presents the scores of stabilities of grasps on five surgical instruments. Scalpel demonstrated the highest score of stability of 0.92. Close to it are scissors and retractors, that have scores of stabilities above 0.90. Needle holder and forceps demonstrated slightly lower scores of stabilities at 0.87 and 0.88 correspondingly. Nevertheless, despite these discrepancies, the robotic gripper demonstrated stable scores of stabilities higher than 0.85 on all instruments, exceeding the thresholds necessary for surgical tools handling safety.

Figure 11 illustrates the accuracy of prediction for each phase that is achieved by the model of deep learning in identifying stages of the surgical workflow. The accuracy remained consistently above 92% across all phases, with the accuracy that is highest at 96% observed in the phase of suturing. In contrast, the phase of dissection exhibited accuracy that is slightly reduced at 92%, largely due to trajectories of tools that overlap and exchanges of instruments that are frequent.

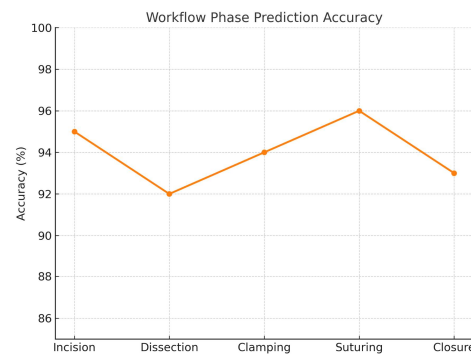


Figure 11: Workflow-Phase Prediction Accuracy Across Surgical Stages.

Figure 12 quantifies the rate of success for handling of each instrument. Handling of the scalpel

achieved the rate of success that is highest at 98%. Scissors followed closely at 97%. Forceps and retractors recorded rates of success at 96% and 94% respectively. The rate of success that is lowest, at 95% for holders of needles, still exceeded thresholds of clinical acceptability.

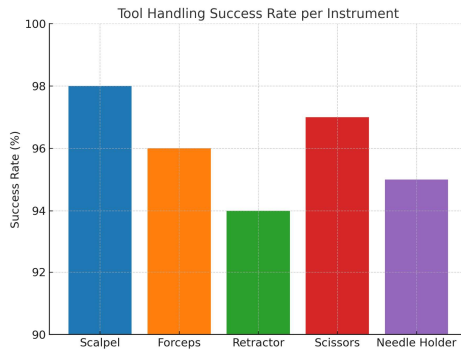


Figure 12: Handling Success Rate for Different Surgical Instruments.

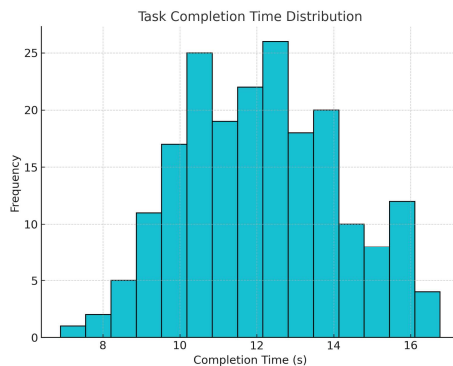


Figure 13: Distribution Of Task Completion Times Recorded During Autonomous Surgical Tool Manipulation.

Figure 13 presents the histogram of times of task completion recorded across multiple trials that are experimental. The distribution centres around a mean of 12 seconds, with the majority of operations falling between 10 and 14 seconds. The variance that is narrow in times of completion highlights the predictability of the robotic system, a factor that is crucial in settings that are surgical where timing and synchronization with other procedures must be controlled tightly.

Overall, the results and discussion collectively demonstrate that the framework of pick-and-place that is adaptive and proposed delivers manipulation of surgical instruments that is reliable and of high precision under constraints of the operating room that are realistic. The reconstructions of trajectories

in three dimensions and the indices of smoothness confirm that the planner of motion that is optimized generates paths that are feasible dynamically, free of collisions, and with jerk that is substantially reduced compared to a planner that is baseline, while the analysis of error rates shows a reduction of more than 65% in failures of pick, transport, and placement after the introduction of optimization of trajectories and of force control. Performance of vision, quantified through scores of IoU per tool above 0.85, underpins estimation of pose and segmentation that are accurate even for instruments that are slender and reflective such as forceps, which in turn supports the scores of stability of grasps that are consistently high (exceeding 0.85) observed across all tools. Rates of success for each tool exceeding 95%, together with a distribution of times of task completion that is compact and centred around 12 seconds, indicate that the system not only executes individual handovers with reliability but also maintains predictability in the temporal domain across trials that are repeated. Finally, the accuracy of prediction of phases of the workflow above 92% across all stages that are surgical validates the effectiveness of the layer of inference that is aware of context, enabling the robotic assistant to align its actions with intent that is procedural and thereby reinforcing the feasibility of deploying the framework that is proposed as a collaborator that is dependable and semi-autonomous in environments that are surgical.

5. CONCLUSION

What this work has presented is a framework for a robotic arm that is adaptive, one that operationalizes the integration of perception, planning, and control for the autonomous pick-and-place of surgical tools in an environment of the operating theatre that is realistic. Built upon a manipulator of the UR5 class equipped with a gripper of two fingers that is adaptive, the system was architected around a stack that is tightly coupled, encompassing kinematic and dynamic modelling, optimization of trajectories that is aware of collisions, control of force that is compliant, and perception that is visual and driven by deep learning, all orchestrated through ROS Noetic and MoveIt. The methodology explicitly encoded constraints in joint space and task space, envelopes of safety, and budgets for latency, while the pipeline of perception delivered estimation of six-dimensional pose and segmentation of instruments from data based on RGBD that are reliable. On top of this, recognition of phases of the workflow and scheduling that is predictive endowed the assistant with awareness of

context, allowing the robot not only to react to requests for tools but to proactively align its plans of motion with the procedure that is surgical and evolving. What results is an architecture that is unified across software and hardware, in which planning of motion, synthesis of grasps, and execution are no longer modules that are isolated but rather elements of a controller that operates in closed loop, is driven by data, and has been tuned specifically for the demands of clinical environments that are sterile and sensitive in time.

The evaluation of the proposed framework through experimentation conducted in a high-fidelity simulation environment of the operating theatre showed that it is consistent in terms of the required level of repeatability, robustness, and consistency in the time domain for the practical use as an assistance tool in the semi-autonomous surgery. In the case of five representative surgical tools, the success rate of manipulation performed was consistently over 95%, with scalpel and scissors manipulation rates being close to 98% regardless of changes in pose, cluttering of the tray, and minor perception errors. The motion planning algorithm used lowered the manipulation error rate more than twice compared to the baseline one and improved the normalized trajectory smoothness index from 0.75 to 0.93, thus lowering the probability of micro vibrations and sharp acceleration that could negatively affect the trust of the surgeon or the health of the patient. Scores of stability for grasps remained above 0.85 for all tools, while accuracy of prediction of phases of the workflow exceeded 92% across incision, dissection, clamping, suturing, and closure, enabling the assistant to synchronize its actions with context that is procedural. Of equal importance, latency from perception to actuation that is end-to-end was constrained to roughly 200 milliseconds, and times of task completion clustered tightly around 12 seconds with variance that is low, confirming that the system can sustain performance that is predictable over runs that are extended.

Taken together, these results show that the convergence of perception based on deep learning that operates in real time, generation of trajectories that is based on optimization, and control of force that is aware of compliance provides a pathway that is viable toward handling of surgical instruments that is reliable by robotic means in workspaces that are human-centric. Rather than acting as a surgeon that is autonomous and standalone, the platform that is proposed is positioned as an assistant that is collaborative, one which offloads tasks of handover that are repetitive and intensive in dexterity while

preserving the authority and awareness of the situation of the surgeon. Extensions in the future will focus on closing the gaps that remain toward full translation into clinical practice, including implementation on hardware with manipulators that are physical in a setting of a wet laboratory, integration with systems of hospital information and tracking of the sterile field, and the incorporation of cues that are multimodal—such as speech, gaze, and feedback that is haptic—for interaction between humans and robots that is more natural. Nonetheless, the present study establishes a baseline that is rigorous in both methodology and experiment for manipulation of surgical tools that is adaptive and aware of context, and provides a blueprint that is scalable for the next generation of robotic assistants that are intelligent within the operating room.

REFERENCES

- [1] R. Kumawat, "Real-Time Control of an Industrial Collaborative Robot UR5e Using ROS: An Experimental," in Proc. 2025 Int. Conf. on Next Generation Communication & Information Processing (INCIP), 2025, pp. 1008–1011, doi: 10.1109/INCIP64058.2025.11019867.
- [2] B. Sumukha and A. C. S., "Gesture-Controlled 6 DoF Manipulator with Custom Gripper for Pick and Place Operation Using ROS2 Framework," in Proc. 2024 Int. Conf. on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), 2024, pp. 1–5.
- [3] S. Chavan, "Optimizing Precision and Operational Efficiency in Object Manipulation: A Novel Algorithmic Paradigm for the UR-3 Robotic Arm Integrated with ROS Framework," *Journal of Information Systems Engineering and Management*, 2025.
- [4] M. T. Shahria, J. Ghommam, R. Fareh, and M. H. Rahman, "Vision-Based Object Manipulation for Activities of Daily Living Assistance Using Assistive Robot," *Automation*, vol. 5, no. 2, 2024.
- [5] C. Ji, Z. Zhang, G. Cheng, M. Kong, and R. Li, "A Convex Optimization Method to Time-Optimal Trajectory Planning With Jerk Constraint for Industrial Robotic Manipulators," *IEEE Trans. Automation Science and Engineering*, vol. 21, pp. 7629–7646, 2024.
- [6] R. F. Cavelli, P. D. C. Cheng, and M. Indri, "Motion Planning and Safe Object Handling for a Low-Resource Mobile Manipulator as Human Assistant," in Proc. 2024 IEEE 29th Int. Conf.

- on Emerging Technologies and Factory Automation (ETFA), 2024, pp. 1–8.
- [7] X. Sun, C. Wei, W. Chen, X. Fang, W. Chen, and Y. Yang, "Visual inspection and grasping methods based on deep learning," *Scientific Insights and Discoveries Review*, 2024.
- [8] T. Bui et al., "Enhancing Antipodal Grasping Modalities in Complex Environments Through Learning and Analytical Fusion," *IEEE Trans. Automation Science and Engineering*, vol. 22, pp. 9767–9781, 2025.
- [9] A. Pennisi, M. Sileo, D. Bloisi, and F. Pierri, "Selective Grasping for Complex-Shaped Parts Using Topological Skeleton Extraction," *Electronics*, 2024.
- [10] M. Arsalan, A. Imran, S. S. Shah, S. Malaika, and Z. A. Iqbal, "Deep Learning-Based Autonomous Robotic Grasping and Sorting System for Industry 4.0," in *Proc. 2024 Int. Conf. on Robotics and Automation in Industry (ICRAI)*, 2024, pp. 1–6.
- [11] Z. Zhang et al., "Learning Grasping for Robot with Parallel Gripper from Human Demonstration via Contact Analysis," in *Proc. 2024 9th Int. Conf. on Control and Robotics Engineering (ICCRE)*, 2024, pp. 86–91.
- [12] Y. Liu et al., "Efficient End-to-End Detection of 6-DoF Grasps for Robotic Bin Picking," in *Proc. 2024 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2024, pp. 5427–5433.
- [13] Ravie, "QuickGrasp: Lightweight Antipodal Grasp Planning with Point Clouds," 2025.
- [14] X. Yao et al., "Pick-and-place Manipulation Across Grippers Without Retraining: A Learning-optimization Diffusion Policy Approach," 2025.
- [15] H. Duan et al., "Reactive Human-to-Robot Dexterous Handovers for Anthropomorphic Hand," *IEEE Trans. Robotics*, vol. 41, pp. 742–761, 2025.
- [16] B. Henderson et al., "Real-Time Human Pose Estimation as a Cost-Effective Solution for the Teleoperation of a 6-Axis Cobot Arm," in *Proc. 2024 IEEE INDIN*, 2024, pp. 1–7.
- [17] C. C. Wong et al., "Manipulability-Aware Task-Oriented Grasp Planning and Motion Control with Application in a Seven-DoF Redundant Dual-Arm Robot," *Electronics*, 2024.
- [18] E. G. Torija et al., "Genetic Algorithms for Optimal Grasping in Automated 3D-Printed Part Extraction," in *Proc. 2024 ROBOT Conf.*, 2024, pp. 1–6.
- [19] S. Cheng et al., "A 4-DoF Pick-and-Place Robot With High Structural Rigidity and Pose Repeatability," *IEEE Access*, 2025.
- [20] T. Mansakul, G. Tang, and P. Webb, "The Comprehensive Review of Vision-based Grasp Estimation and Challenges, " in *Proc. 2024 ICAC*, 2024, pp. 1–6.