Teaching Large Language Models to Create Task Plans for Robotic Suturing

Megan Zhou[1] and Mahdi Tavakoli[1,2]
[1]Department of Biomedical Engineering, University of Alberta, Edmonton, Canada.
[2]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada.
Email: jzhou17@ualberta.ca

INTRODUCTION

Robot-assisted surgery can improve surgical precision and potentially reduce errors compared to purely manual procedures [1]. Autonomy in surgical robots requires generating high-level task plans that can be executed reliably in complex, dynamic environments like those involving deformable tissue [2]. Large Language Models (LLMs) can assist by interpreting procedural descriptions and generating structured, machine-readable plans. We investigated the ability of an LLM to generate Planning Domain Definition Language (PDDL) task descriptions for robotic suturing, which can be integrated into task and motion planning pipelines (Fig 1).

MATERIALS AND METHODS

We designed a workflow in which surgical procedure description texts were provided to an LLM (ChatGPT 5). The model was instructed to output two PDDL files: a *domain file* describing the physical environment and actions, and a *problem file* specifying initial and goal states. Generated files were then uploaded to PDDL Editor [3] to check the syntax. Upon receiving syntax errors, we prompted ChatGPT to fix them. Once PDDL Editor reported no errors, we passed the files to Fast Downward task planner [4] and analyzed the outputted action sequence for completeness and accuracy according to the procedure descriptions. We evaluated the LLM's ability to generate syntactically correct PDDL files that produced logical action sequences for the suturing task in [5].

RESULTS AND DISCUSSION

Our initial experiments on robotic suturing yielded syntactically correct PDDL files in 4 out of 5 trials without manual intervention. In the first trial, syntax errors were unbalanced parentheses in the domain file and illegal state logic in the problem file. We prompted ChatGPT to fix these errors, resulting in files that were syntactically valid. 4 out of 5 Fast Downward task plans were accurate to suturing steps in [5], yielding action sequences like (reach-for needle), (position-needle-tip), and (push-needle-through-tissue). The fourth trial presented a logical error in which a step near the end of the procedure appeared as the second action in the task plan. After pointing out this error to ChatGPT, it created files that produced the correct step order. These results suggest that LLMs can reliably produce machine readable task descriptions from natural language instructions, but prompt refinement is critical to minimize logic and syntax errors.

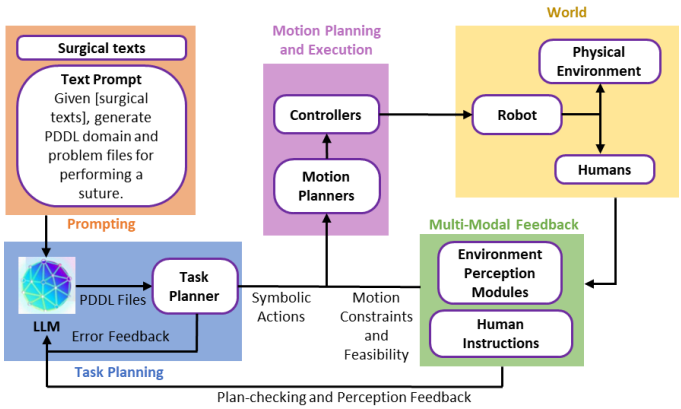

Fig 1 System components.

CONCLUSIONS

This study shows that LLMs can generate valid PDDL task descriptions for robotic suturing, enabling task planners to produce symbolic action sequences. Future work will focus on improving prompt robustness, quantifying PDDL file generation accuracy over a larger set of surgical tasks, and integrating the generated plans into an autonomous robotic execution pipeline.

REFERENCES

[1] A. Handa et al., "Role of Robotic-Assisted Surgery in Public Health: Its Advantages and Challenges," in Cureus, 2024.
[2] A. Shademan et al., "Supervised autonomous robotic soft tissue surgery," in Science Translational Medicine, vol. 8, no. 337, 2016.
[3] PDDL Editor, https://editor.planning.domains
[4] Fast Downward, https://www.fast-downward.org
[5] N. Ahmidi et al., "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 9, pp. 2025–2041, 2017.