Improving Technical Proficiency in Robot-mediated Surgery Through Counterfactual Inquiry

Project Proposal

EN 601.656 Computer Integrated Surgery II

Hao Ding

hding15@jhu.edu

Project Description & Goal

The quality of a robot-mediated surgery is highly related to the skill of the surgeon. The more skillful the surgeon is, the better the patient output will be. Thus, improving technical proficiency is always worth for researchers to look into. Our goal for this project is to develop a robust and interpretable system to empower the novice surgeons. One way to achieve this is to translate the novice commands to the surgery robot into the commands that are more likely to be given by a more experienced and skillful expert surgeons.

Recent deep learning algorithms dominate large amounts of benchmarks for some areas including action recognition and prediction. Deep learning's success in the action prediction provides us a powerful method for the surgery assistance. We may treat the assistance as a kinematic prediction task using a deep neural network with the provided context and surgery task as input. Empirically expected, if designed and implemented properly and with enough training data provided, the deep learning methods will produce the state-of-the-art performance to assist a novice surgeon during surgery. However, deep learning methods have their drawbacks. The most significant and related drawbacks are generalization issue and low interpretability. If the environment of the surgery is not from the same distribution where the training data of the networks are generated, the deep learning methods have large probabilities to fail in incomprehensible ways. These drawbacks make the assistance system not reliable enough to put into practice.

Causal inference mechanisms, with causal relation provided, are recently being highly focused on and explored by researchers to deal with the generalization issues and low interpretability of the deep learning algorithms. We will follow some recent works in incorporating causal inference mechanisms into deep learning architectures and try to improve the robustness of the assistance system through counterfactual inquiry.

Technical Summary of Approach

The overall approach of building a robust system to assist surgeons especially novice surgeons perform at expert-level by a simple query "Do something (task) somewhere (task context) as an expert" has 3 aspects – data, human understandings and algorithms.

As the figure 1 shown, data is the prerequisite of both understandings and algorithms. We need to understand the difference of commands given by novice and expert surgeons by some statistical analysis for the data or some intuitional feelings after viewing sufficient videos of surgery operations. At the same time, we need data to train the network. And human understanding of the surgery procedure will help in designing both the deep learning algorithms and causal inference algorithms.

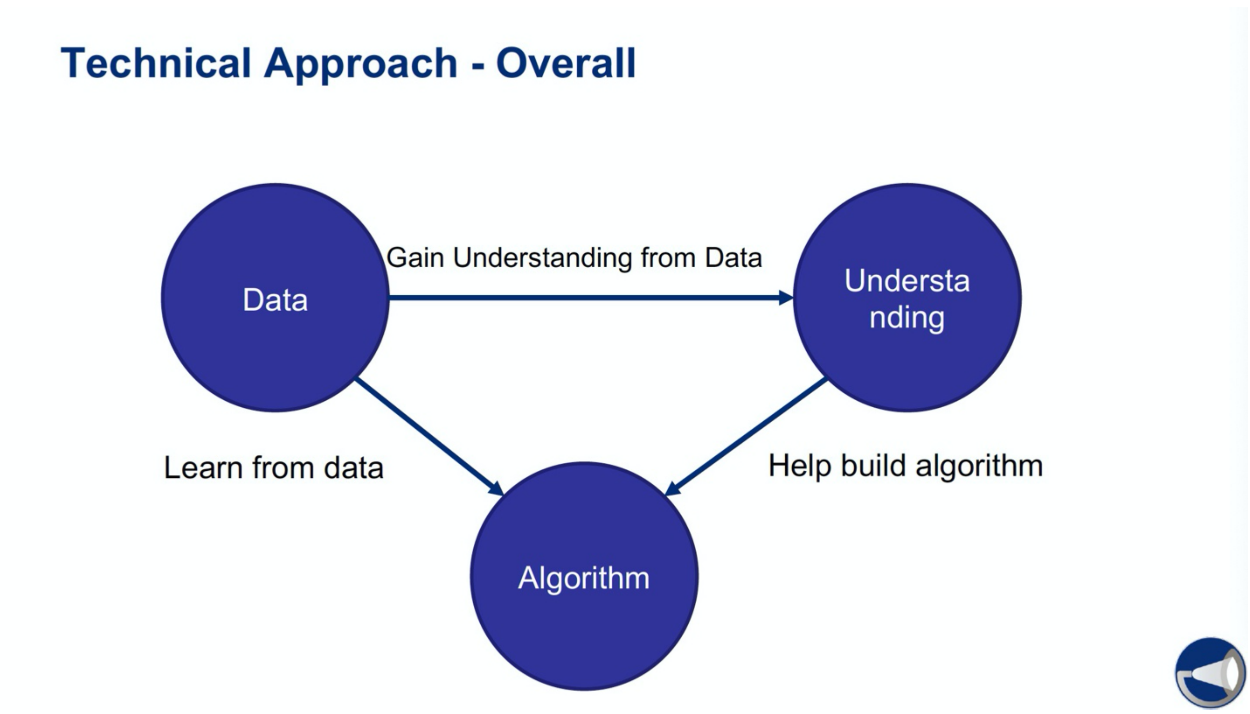

Figure 1. Overall technical approach

Technical Summary of Approach

Data

The foundation of our data is the JHU-ISI Gesture and Skill Assessment Working Set (JIGSAW) which was captured using the da Vinci Surgical System from eight surgeons with different levels of skill performing five repetitions of three elementary surgical tasks on a bench-top model. The JIGSAW dataset contains 3 components: kinematic data, video data and human annotation of skill level and gestures. The figure 2 are examples of video frames in JIGSAW dataset on different tasks (suturing, knot tying, needle passing)

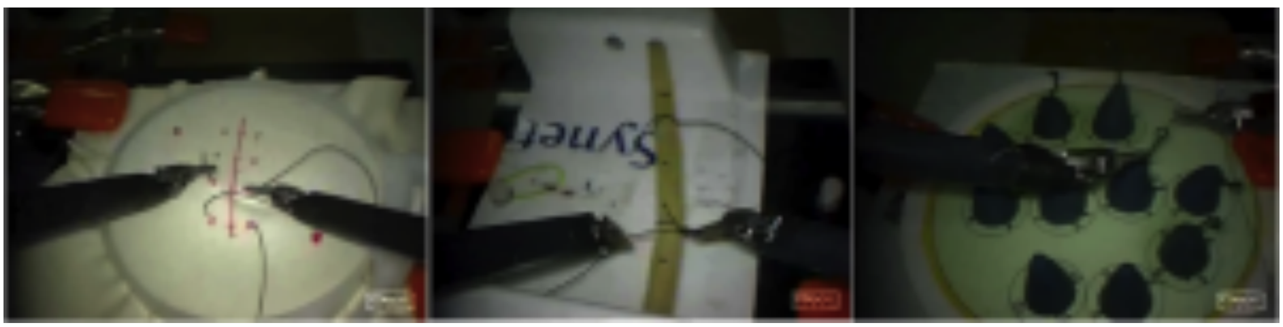

Figure 2. examples of video frames in JIGSAW

Our target on the dataset is to temporally align the video samples for causal analysis. To achieve this, we will firstly manually segment and pair videos and then make paired segments into same length. We will use dynamic time warping algorithm with the l1 norm between normalized traveled distance as the distance function to align the video segments of different length.

Understanding

To perceive more deep understanding of the surgery procedure and the difference between operation given by novice and expert surgeons, we will perform some statistical analysis for properties like path length, time consumed, number of movements of the trajectories of the robot motion.

Algorithms

The algorithms have two parts, deep learning part and causal inference part.

For the deep learning part, our design is based on a basic transformer network. The input to the network will be: (1) Task information: task id and its context (2) Expected skill level: novice- or expert- level (3) Kinematics of previous frames. The first two will be input into the encoder part and the third one will be input into the decoder. The output of the network will be the kinematics of the following frames.

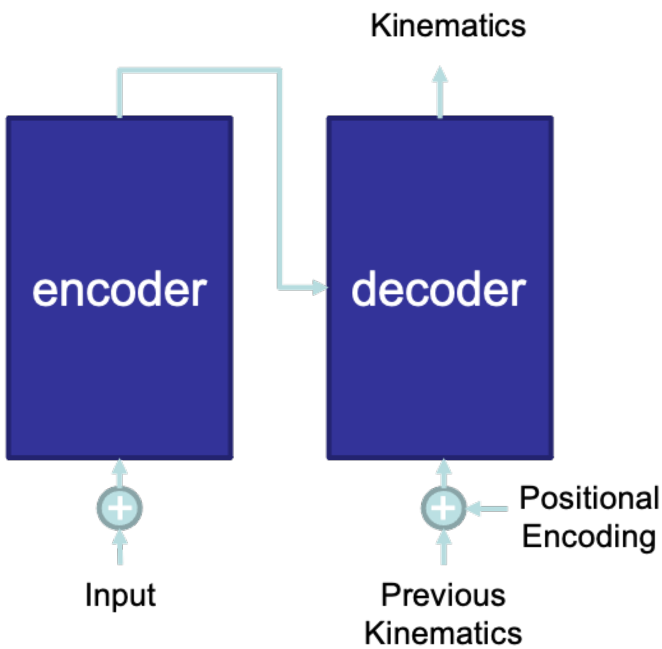

Figure 3. Illustration of the network

For the causal inference part, the causal model is shown in the figure 4, where we have 4 nodes, task id, context, skill level and kinematic sequence. The context node is one of the direct causes of the choice of the task. Then the context and task choice and the skill level are all direct cause of the kinematic sequence.

To better describe the context and the task, besides feed in the video frame we might also extract the context affordance of the context like for suturing task we will extract the in state and the out state from the video frame of the phantom.

As for how to incorporate the causal inference mechanism of this causal model into the deep learning architecture we will refer to the following works to find out an effective way.

- Deep Structural Causal Models for Tractable Counterfactual Inference https://arxiv.org/abs/2006.06485
- CausalGAN: Learning Causal Implicit Generative Models with Adversarial Training https://arxiv.org/pdf/1709.02023.pdf
- Learning Functional Causal Models with Generative Neural Networks https://arxiv.org/abs/1709.05321
- Counterfactual Generative Networks https://arxiv.org/pdf/2101.06046v1.pdf

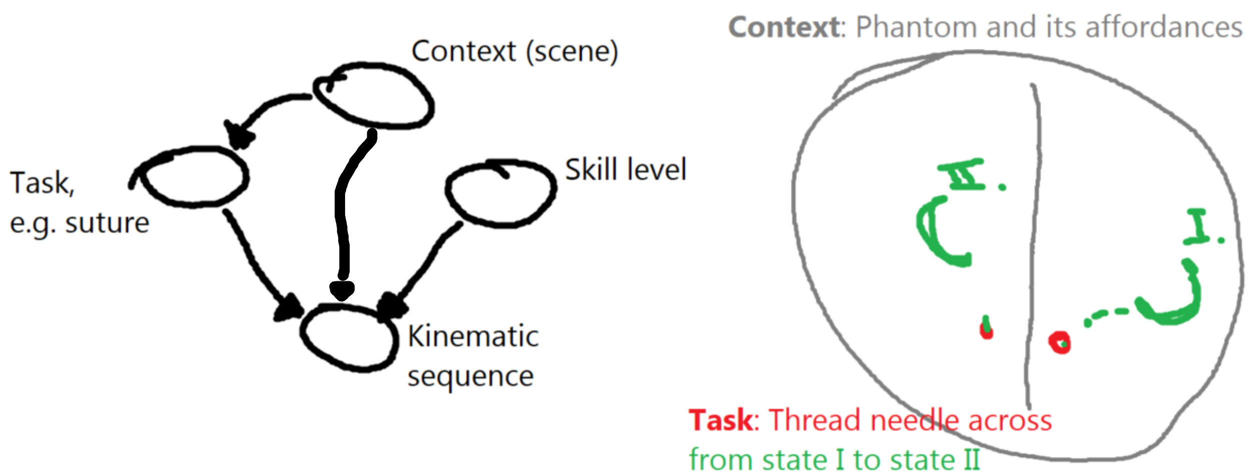

Figure 4. Causal Model and Illustration of Context

Deliverables

Level	Activities	Deliverables
Minimal	Manually segment and pair videos.	Task-aligned JIGSAW dataset for causal analysis
	Implement DTW to align videos segments temporally	
Expected	Statistical analysis for some properties of the trajectory.	Written analysis of understandings into novice- and expert-level robot command
Stretch	Implement the Kinematic Prediction Network	Development and evaluation of counterfactual model
	Incorporate Causal inference mechanism into the DL methods	

Table 1. Deliverables

The minimum target for this semester is to finish the preparation of the datasets. The dataset is the prerequisite both for the analysis and the algorithm design. This task will be performed by two activities – firstly manually segment and pair the videos, afterwards implement the dynamic time warping algorithm to temporally align those pairs of videos. After all the activities are done, a task-aligned JIGSAW dataset will be provided.

The expected target is to make a statistical analysis for the properties of the moving trajectory of the operations in the prepared dataset. A written analysis of the understandings into the novice- and expert-level robot command will be provided.

The Stretch target is to implement the algorithms for the kinematic predictions. Firstly, a deep learning algorithm will be implemented along with the experiments of the performance on the prepared dataset will be operated. Then the causal inference mechanism will be incorporated into the deep learning algorithm and the experiments will be operated on the same dataset. The code and the report for the experiments will be provided.

Timeline

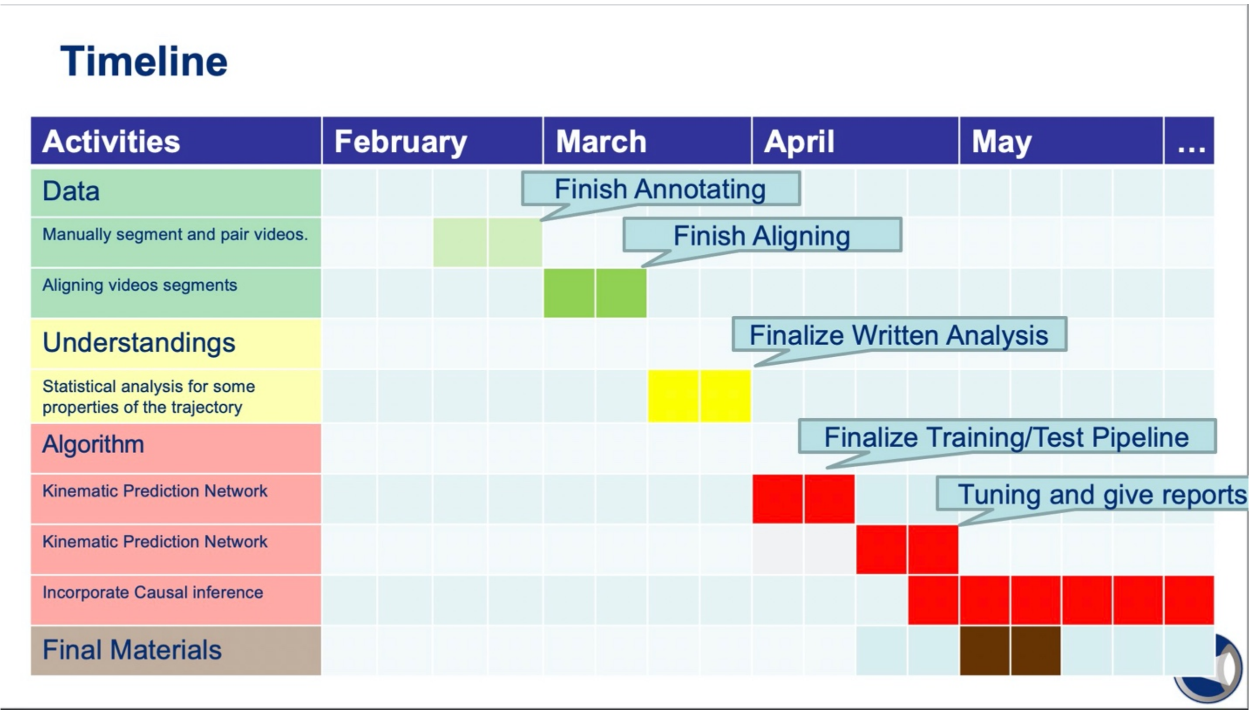

Figure 5. Gantt Chart

Above is the timeline for the project. The data part has the first priority, the manual segmentation and paring will be start at the middle of the February and will be finished before March, after that the implementation of the aligning algorithm will be start and will be finished at the middle March.

Then with the paired the data, the statistical analysis will be started immediately and will be finished in two weeks.

At the start of April, the activities for the algorithms will be launched and hopefully the pure deep learning part will be finished in 1 month. This will be divided into 2 procedures. The first is to implement the training and testing pipelines, which is planned to be done in the first two weeks. Then tuning the networks and perform the experiments will be done in the following two weeks. During experiments, the design and implementation of the causal inference empowered algorithm will be started.

The preparation of the final materials for the course will be started at the head of May.

Dependencies

Dependency	Need	Status	Contingency Plan
JIGSAW dataset	Fundamental dataset	Acquired	N/A
Computational Resources	For Deep Learning experiments	Acquired	If crashed, ask Dr. Unberath for other computers, or acquire on some clouds resources (e.g. AWS).

Table 2. Dependencies

Since our projects is performed purely on public datasets and is more related to analysis and algorithms, our dependencies are not difficult to get. The JIGSAW dataset has been downloaded from its official website and don't have any contingency to be considered. LCSR has provided thin6 server with 3 GPUs as the computational resources for the project. If that sever crashed down, I will contact Dr. Unberath for other computational resources or try to acquire from some cloud server providers (e.g. AWS)

Team Members

- Hao Ding (hding15)
  Intro: *First-year Ph.D. student, Whiting school of engineering, Computer Science Department.*
  Responsibility: *Sole responsibility for all tasks*

Mentors

- Dr. Mathias Unberath
  Intro*: Assistant Professor****,*** *Whiting school of engineering Computer Science Department.*
  Responsibility: *Advising students for all tasks.*

Management Plan

- **Weekly meeting:**
  Weekly meeting with Robotics group in ARCADE Lab including Dr. Mathias Unberath.
- **Biweekly meeting:**
  Biweekly personal meeting with Dr. Mathias Unberath.

Reading list

- Murat Kocaoglu, Christopher Snyder, Alexandros G. Dimakis, and Sriram Vishwanath. Causalgan: Learning causal implicit generative models with adversarial training. In 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018
- hristos Louizos, Uri Shalit, Joris M. Mooij, David A. Sontag, Richard S. Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pages 6446–6456, 2017
- Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.
- Yu Kong and Yun Fu. Human action recognition and prediction: A survey. CoRR, abs/1806.11230, 2018

References

- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pages 5998–6008, 2017.

- Yixin Gao, S. Swaroop Vedula, Carol E. Reiley, Narges Ahmidi, Balakrishnan Varadarajan, Henry C. Lin, Lingling Tao, Luca Zappella, Benjam´ın B´ejar, David D. Yuh, Chi Chiung Grace Chen, Ren´e Vidal, Sanjeev Khudanpur and Gregory D. Hager, *The JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling*, In Modeling and Monitoring of Computer Assisted Interventions (M2CAI) – MICCAI Workshop, 2014.

- Narges Ahmidi, Lingling Tao, Shahin Sefati, Yixin Gao, Colin Lea, Benjamın Bejar Haro, Luca Zappella, Sanjeev Khudanpur, Rene Vidal, Fellow, IEEE, Gregory D. Hager, *A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery,* Transaction of Biomedical Engineering, 2017.