B24 (Fri-AM)

MICCAI2025 Daejeon REPUBLIC OF KOREA

MedAIR @CUHK

Surgical Action Planning with Large Language Models

Mengya Xu[1*], Zhongzhen Huang[2,3*], Jie Zhang[4],
Xiaofan Zhang[2,3], and Qi Dou[1 ✉]

*: Equal technical contribution, ✉: Corresponding author
[1] The Chinese University of Hong Kong, Hong Kong SAR, China
[2] Shanghai Jiao Tong University, Shanghai, China
[3] Shanghai AI Laboratory, Shanghai, China
[4] Huazhong University of Science and Technology, Wuhan, China

Highlights

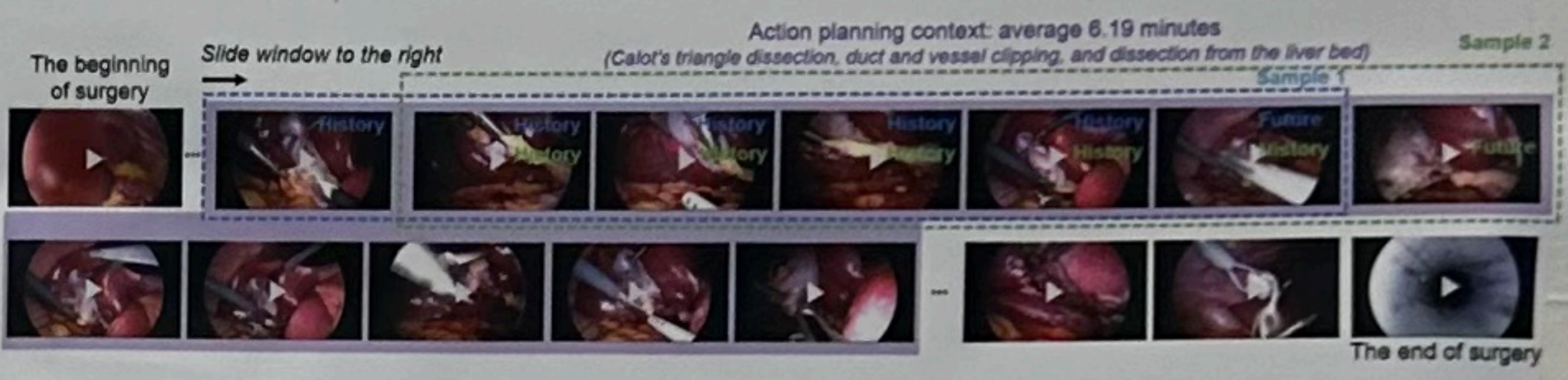
□ Introduction of the Surgical Action Planning (SAP) task in RMIS for generating future surgical action plans from visual inputs.
□ Development of the LLM-SAP framework with Near-History Focus Memory (NHFM) module and prompts factory for action prediction.
□ Flexible solution for zero-shot and fine-tuning customization of LLMs and VLMs through visual observation processing.
□ Evaluation of LLM-SAP on CholecT50-SAP dataset using SOTA models (Qwen2.5, Qwen2-VL) in zero-shot and fine-tuning settings.

Surgical Action Planning Task Definition

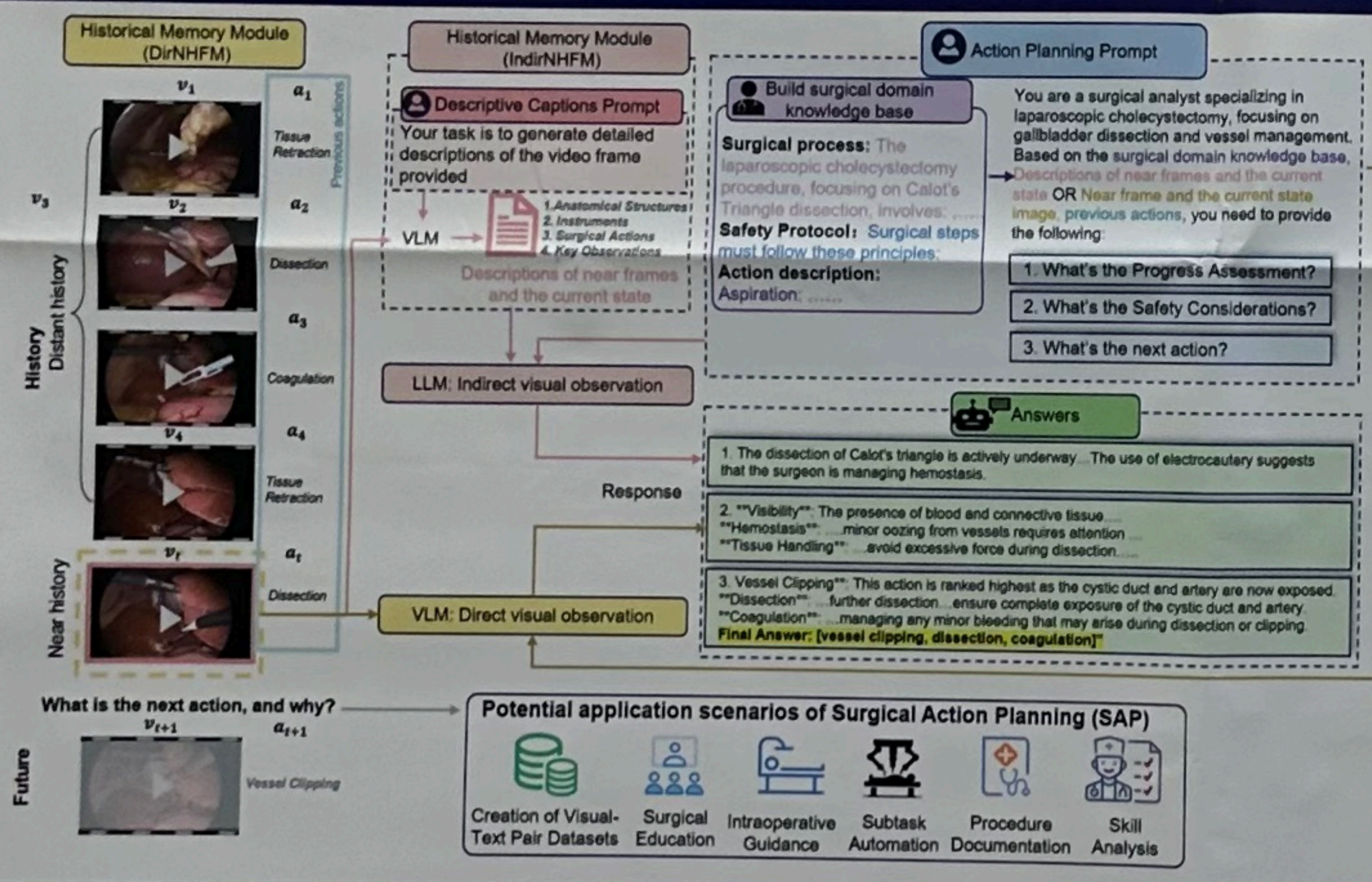
In SAP, the model generates an action plan A = {a1, . . . , at} by leveraging two key inputs: a visual history H and a user-defined goal G, aiming to transition the current state to the desired goal within a planning horizon of T steps. The visual history H, represented as a sequence of video clips {v1, . . . , vt}, captures the progression toward the goal over time. Meanwhile, the goal G is expressed as a natural language description, such as "Provide analysis and the next action for laparoscopic cholecystectomy." Each action at in the plan corresponds to a categorical label within a set of C possible actions.

Dataset Construction

□ Group consecutive frames sharing the same action into action clips. The action labels include Aspiration, Coagulation, Dissection, Tissue Retraction, Vessel Clipping.
□ Focus on the video segments: Calot's triangle dissection, duct and vessel clipping, and dissection from the liver bed.
□ The 50 chosen video segments have an average duration of 6.19 minutes each.
□ One Sample = 5 historical action clips, and the subsequent 1 action (future one-step).
□ These 50 video segments comprise a total of 225 samples. Training: 35 videos, 168 samples; Testing:15 videos, 57 samples.


The beginning of surgery
Slide window to the right
Action planning context: average 6.19 minutes
(Calot's triangle dissection, duct and vessel clipping, and dissection from the liver bed)
Sample 2
Sample 1
History
Future
The end of surgery

Method


Historical Memory Module (DirNHFM)
Historical Memory Module (IndirNHFM)
Descriptive Captions Prompt
Your task is to generate detailed descriptions of the video frame provided
VLM
1. Anatomical Structures
2. Instruments
3. Surgical Actions
4. Key Observations
Descriptions of near frames and the current state
LLM: Indirect visual observation
VLM: Direct visual observation
Response
Action Planning Prompt
Build surgical domain knowledge base
Surgical process: The laparoscopic cholecystectomy procedure, focusing on Calot's Triangle dissection, involves: ......
Safety Protocol: Surgical steps must follow these principles:
Action description:
Aspiration: ......
You are a surgical analyst specializing in laparoscopic cholecystectomy, focusing on gallbladder dissection and vessel management. Based on the surgical domain knowledge base, Descriptions of near frames and the current state OR Near frame and the current state image, previous actions, you need to provide the following:
1. What's the Progress Assessment?
2. What's the Safety Considerations?
3. What's the next action?
Answers
1. The dissection of Calot's triangle is actively underway... The use of electrocautery suggests that the surgeon is managing hemostasis.
2. **Visibility**: The presence of blood and connective tissue... **Hemostasis**: ...minor oozing from vessels requires attention... **Tissue Handling**: ...avoid excessive force during dissection...
3. Vessel Clipping**: This action is ranked highest as the cystic duct and artery are now exposed. **Dissection**: ...further dissection...ensure complete exposure of the cystic duct and artery. **Coagulation**: ...managing any minor bleeding that may arise during dissection or clipping.
Final Answer: [vessel clipping, dissection, coagulation]"
What is the next action, and why?
Potential application scenarios of Surgical Action Planning (SAP)
Creation of Visual-Text Pair Datasets
Surgical Education
Intraoperative Guidance
Subtask Automation
Procedure Documentation
Skill Analysis

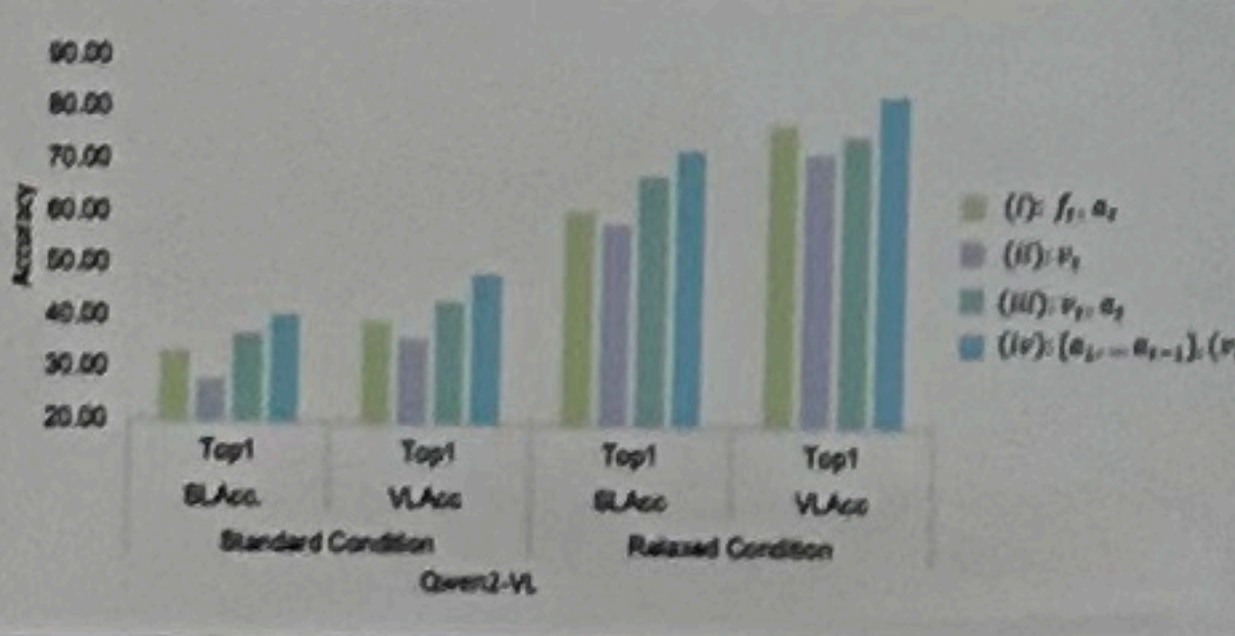
- **LLM-SAP architecture** for predicting next surgical actions and forming long-horizon action chains using advanced LLM (Qwen2.5) and VLM (Qwen2-VL).
- **Two versions:** Text-based LLM planning (IndirNHFM) using text descriptions of visual history; VLM planning (DirNHFM) using visual history directly.
- **Near-History Focus Memory Module (NHFM):** emphasizes a compact summary of past actions for the distant history while focusing on detailed information from the near history. The "near" history refers to the most recent historical action clip (frames) immediately preceding the current observation. The "distant" history comprises all historical action labels (text) that precede the 'near' history.

$$\text{VLM}(\{a_i \mid i == 1, \ldots, t-1\}, \langle v_t, a_t \rangle, \text{Prompt})$$

- **Action planning prompts**
➢ **System prompts:** Surgical Domain Knowledge Base is build based on surgical process, safety protocol, and action description.
➢ **User prompts:** structed outputs including progress assessment, safety considerations, and future action recommendations.
➢ **Descriptive captions prompts:** To generate descriptive captions for frames.

Quantitative Results

Models	HMM	Standard Condition						Relaxed Condition					
		SLAcc			VLAcc			Re SLAcc			Re VLAcc		
		Top1	Top2	Top3	Top1	Top2	Top3	Top1	Top2	Top3	Top1	Top2	Top3
Zero-Shot													
Qwen2.5-32B	IndirNHFM	45.61	63.16	66.67	**53.42**	62.48	64.65	67.44	95.35	**97.67**	78.97	**97.38**	98.33
Qwen2.5-72B		45.61	59.65	66.67	**53.42**	59.31	64.65	67.44	83.72	**97.67**	78.97	80.21	**99.05**
Qwen2-VL	DirNHFM	40.35	57.89	68.42	48.37	56.37	66.32	**72.09**	88.37	90.70	**82.48**	93.45	95.24
Supervised Fine-Tuning													
Qwen2.5-72B-SFT	IndirNHFM	45.61	**78.95**	80.70	**53.42**	80.05	82.27	67.44	93.02	**97.67**	78.97	93.33	98.33
Qwen2.5-32B-SFT		45.61	**78.95**	**85.96**	**53.42**	**83.05**	**87.89**	67.44	**95.35**	**97.67**	78.97	96.67	98.33
Qwen2-VL-72B-SFT	DirNHFM	**47.37**	54.39	68.42	**54.31**	58.55	65.98	60.47	74.42	90.70	67.30	83.25	94.76



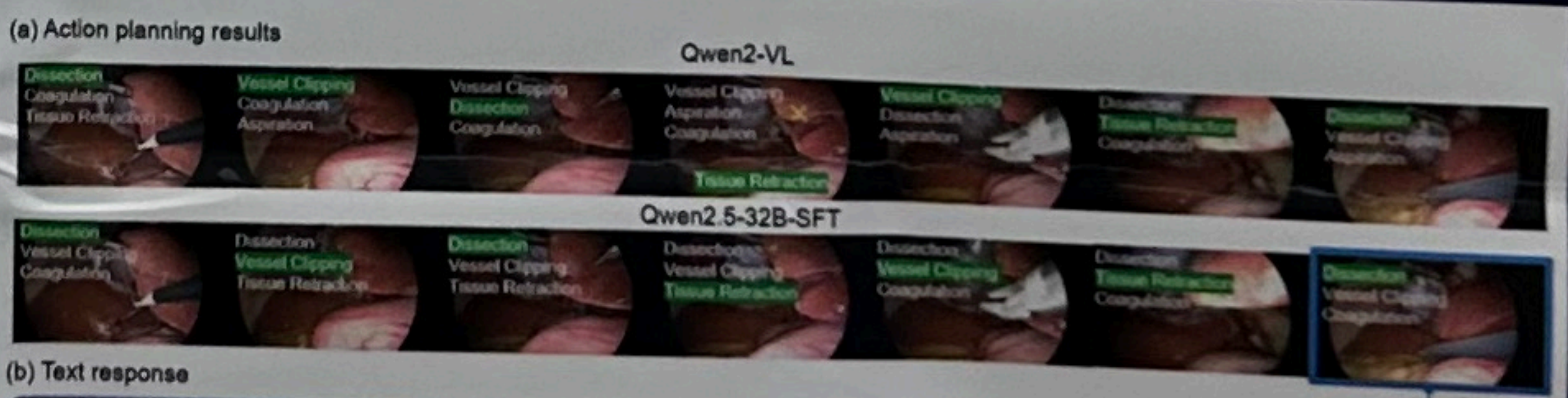
- **SLAcc:** the predicted next action $\hat{a}_i$ to exactly match the ground truth action a_i
- **VLAcc:** evaluate the model's performance on individual surgical patients
- **ReAcc:** the model's suggested actions $\hat{a}_i$ occur within the current step or the subsequent one-step ($\hat{a}_i \in \{a_i, a_{i+1}\}$)

Conclusion

•**Introduction of SAP Task:** Proposing the Surgical Action Planning (SAP) task for forward-looking decision-making in surgery, addressing challenges like instrument-action complexity, temporal dependencies, and data privacy.
•**LLM-SAP Framework:** A novel solution leveraging LLMs with NHF-MM and prompts factory to predict actions and provide interpretable responses, validated on our constructed CholecT50-SAP dataset with SOTA models (Qwen2.5/QwenVL).

Qualitative Results

(a) Action planning results



(b) Text response

Progress Assessment
The current surgical progress indicates that the gallbladder is being effectively retracted to enhance visibility of Calot's triangle. However, the critical structures, namely the cystic duct and cystic artery, remain obscured by connective tissue and fat. The previous actions suggest a focus on vessel clipping and dissection, but the last action of tissue retraction indicates that the procedure is still in the preparatory phase for dissection. Overall, the visibility is acceptable, but further dissection is necessary to achieve the Critical View of Safety (CVS).

Safety Considerations
There are several safety considerations at this stage:
1. **Visibility**: The obscured view of the cystic duct and artery poses a risk of inadvertent injury to surrounding structures if dissection is not performed carefully.
2. **Tissue Handling**: Care must be taken to avoid excessive force during dissection to prevent tearing or bleeding, especially near vascular structures.
3. **Hemostasis**: Although there is currently no evidence of bleeding, continuous monitoring for hemostasis is essential as dissection progresses.

Ready-to-Execute Actions
1. **Dissection**: This action is ranked highest as it is necessary to clear the connective tissue and fat obscuring Calot's triangle. The current state indicates readiness for this step, as the gallbladder is already retracted, providing a clear path for dissection.
2. **Vessel Clipping**: This action is ranked second because once the cystic duct and artery are identified through dissection, securing them with clips will be crucial. However, it cannot be executed until the structures are adequately exposed.
3. **Coagulation**: This action is ranked third as it may be required during dissection to control any minor bleeding that could arise from the manipulation of tissues. While important, it is contingent upon first achieving a clear view of the anatomical structures.

Final Answer [Dissection, Vessel Clipping, Coagulation]