Internship Offer – Master 2 or Engineering Level

Uncertainty Quantification for Action Triplet Prediction in Surgical Robotics

Dates: From Spring 2026, 6 months
Location: Sorbonne Université, ISIR
Contacts: mohamed.chetouani@isir.upmc.fr, nicolas.thome@isir.upmc.fr, rambour@isir.upmc.fr

Funding is available to pursue this topic through a PhD.

Summary

Modern surgical robotics relies increasingly on AI-based perception to recognize instruments, actions, and anatomical targets during procedures. A compact and informative way to describe such interactions is through action triplets ⟨instrument, verb, target⟩ — for example, ⟨Grasper, Grasp, Gallbladder⟩.

Recent Vision–Language Models (VLMs) show promising capabilities for modeling these complex interactions. However, a key challenge remains: enabling such models to quantify their uncertainty, recognize ambiguous or risky situations, and signal when predictions may be unreliable.

This internship focuses on developing uncertainty quantification (UQ) methods for action triplet prediction in surgical scenes, with the long-term goal of improving trust, safety, and interpretability in robot-assisted spine surgery.

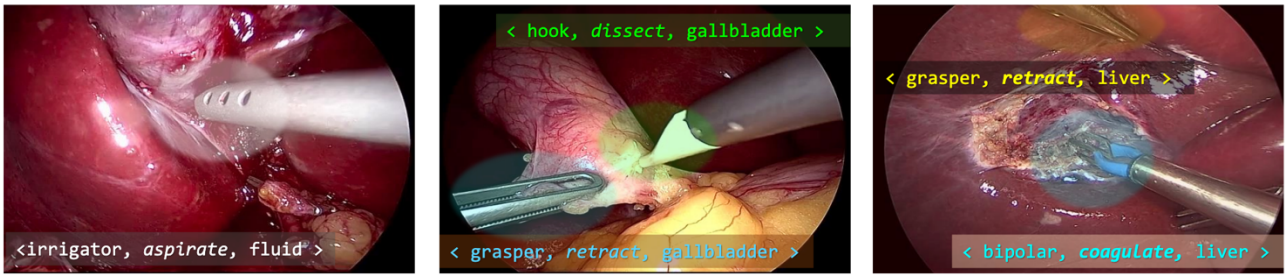

Figure 1: *The goal of the internship is to provide reliable Uncertainty Quantification (UQ) for action triplet predictions in surgical robotics. Ultimately, UQ can be leveraged to design advanced human-robot interaction schemes (figure from [5]).*

Objectives

The goal of the internship is to develop and evaluate **UQ strategies** for VLM-based action triplet recognition, and to explore their transfer to spine surgery.

Uncertainty Quantification for Action Triplets

- Design methods to estimate **aleatoric** (data) [1,2], **epistemic** (model) [3], and **semantic** uncertainties [4] in triplet predictions.
- Detect ambiguous or out-of-distribution cases (e.g., unseen instruments, anatomy, or lighting conditions).
- Identify and interpret **semantic misalignment errors**, where the model predicts a plausible but incorrect triplet (e.g., ⟨Scissors, Cut, Cystic duct⟩ instead of ⟨Clipper, Clip, Cystic duct⟩).

Evaluation on Surgical Datasets

- **CholecTriplet2021 (CholecT50) [5]**: 50 annotated laparoscopic cholecystectomy videos (≈100k frames, 161k triplets) with binary presence labels for 100 action triplet classes (6 instruments, 10 verbs, 15 targets).
- **CholecTriplet-Seg [6]**: >30,000 annotated frames linking instrument instance masks with verbs and anatomical targets, enabling instance-level triplet grounding and strong supervision.
- **CholecTrack20 [7]**: 20 laparoscopic cholecystectomy videos with rich annotations for **multi-class multi-tool tracking**, **phase recognition**, and **scene-level visual challenges**. This dataset provides **multi-perspective trajectories** for tool visibility and movement, enabling temporal uncertainty analysis.

Transfer to Spine Surgery

- Adapt and apply the developed methods to a **custom spine surgery dataset** within the RODEO robotic framework.
- Explore **uncertainty-informed feedback** for human–robot collaboration, where the system can explain its confidence ("Segmentation uncertain due to occlusion") or request clarification ("Move camera to improve visibility").

[1] Confidence Estimation via Auxiliary Models. C. Corbière, N. Thome, A. Saporta, T-H. Vu, M. Cord, P. Pérez. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[2] ViLU: Learning Vision-Language Uncertainties for Failure Prediction. M. Lafon, Y. Karmim, J. Silva-Rodriguez, P. Couairon, C. Rambour, R. Fournier, I. Ben Ayed, J. Dolz, N. Thome. ICCV 2025.

[3] Hybrid Energy Based Model in the Feature Space for Out-of-Distribution Detection. M. Lafon, E. Ramzi, C. Rambour, N. Thome. ICML 2023.

[4] I-FailSense: Towards General Robotic Failure Detection with Vision-Language Models. C. Grislain, H. Rahimi, O. Sigaud, M. Chetouani. Arxiv, 2025.

[5] Chinedu Innocent Nwoye, et. al. CholecTriplet2021: A benchmark challenge for surgical action triplet recognition. Medical Image Analysis, Volume 86, 2023.

[6] Grounding Surgical Action Triplets with Instrument Instance Segmentation: A Dataset and Target-Aware Fusion Approach. O Alabi, M Wei, C Budd, T Vercauteren, M Shi. ArXiv, 2025.
[7] A Multi-Perspective Tracking Dataset for Surgical Tools. C.I. Nwoye, K. Elgohary, A. Srinivas, F. Zaid, J. L. Lavanchy, N. Padoy. CVPR 2025.