Semi-Supervised Surgical Tool Detection

DeAndre Bailey |Mentor: *Darinka Townsend Hinostroza* | Mentor: *Yuanhao Zou* | Advisor: *Dr. Chen Chen*

UCF | Institute of Artificial Intelligence

Project Introduction

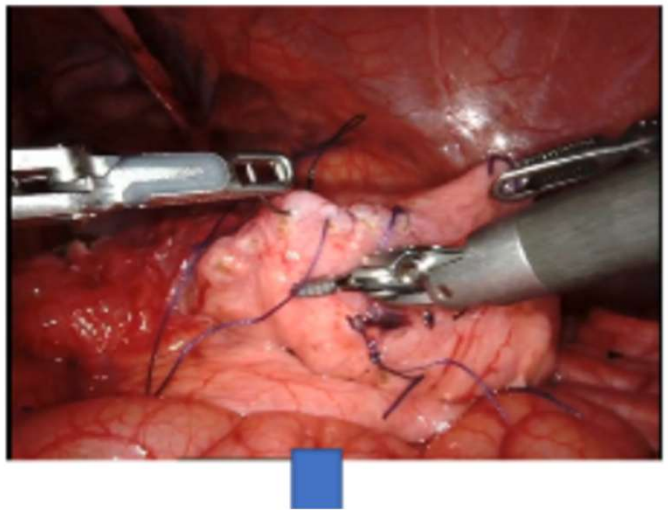
- Surgical Visual Understanding 2026 Category 1 Challenge
- Classify and localize surgical instruments using bounding boxes
- Develop **semi/weakly supervised architecture**



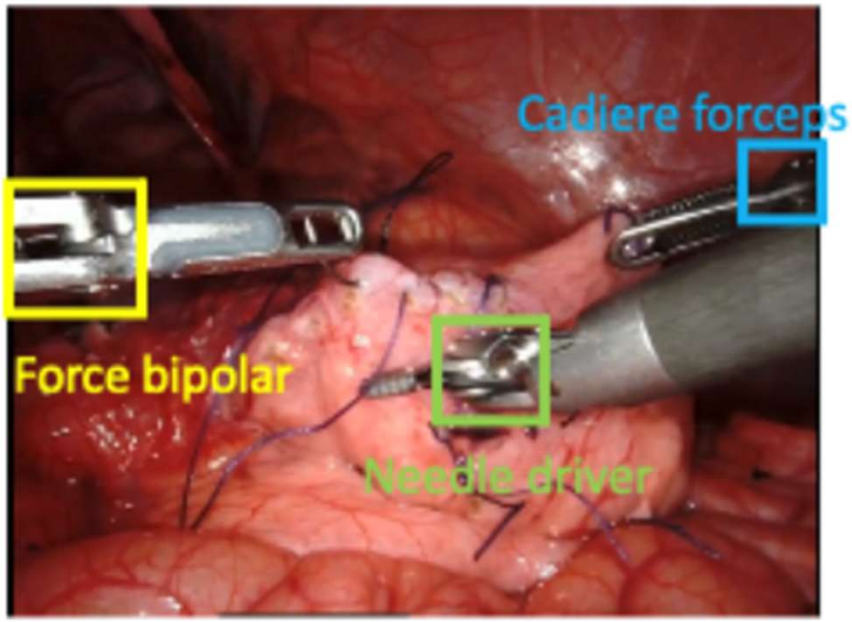
Project Introduction

Category 1
(weak/noisy labels)

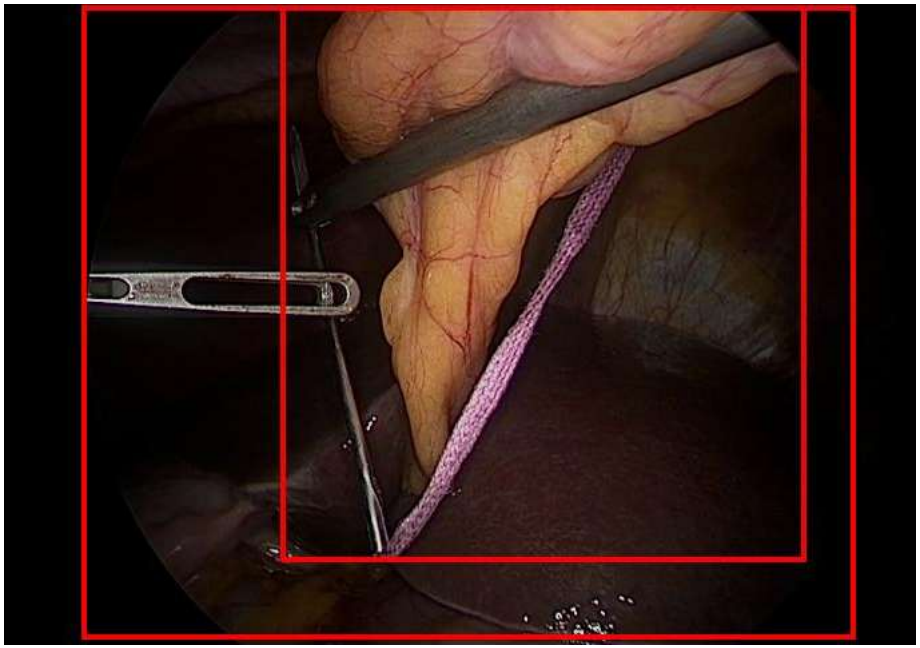
Input image



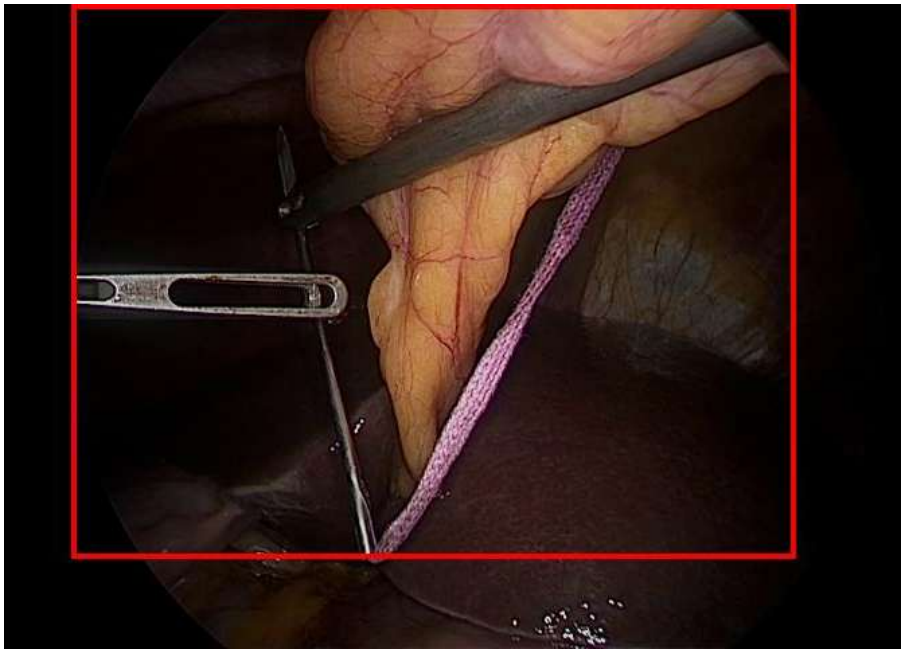
Classify and localize tools



Domain Gap



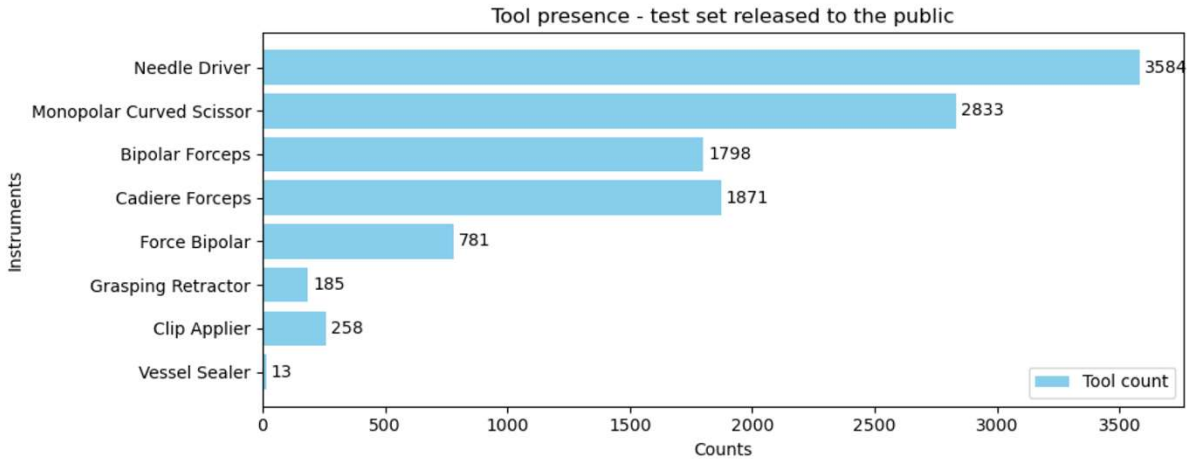
LocateAnything Output Example



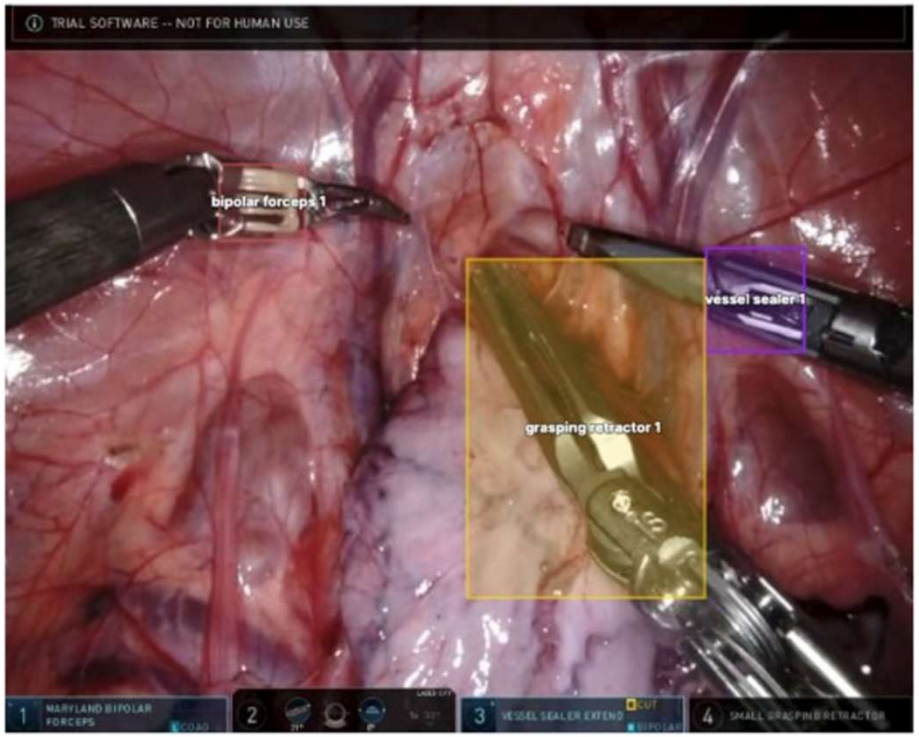
Grounding Dino Output Example

Data Gap

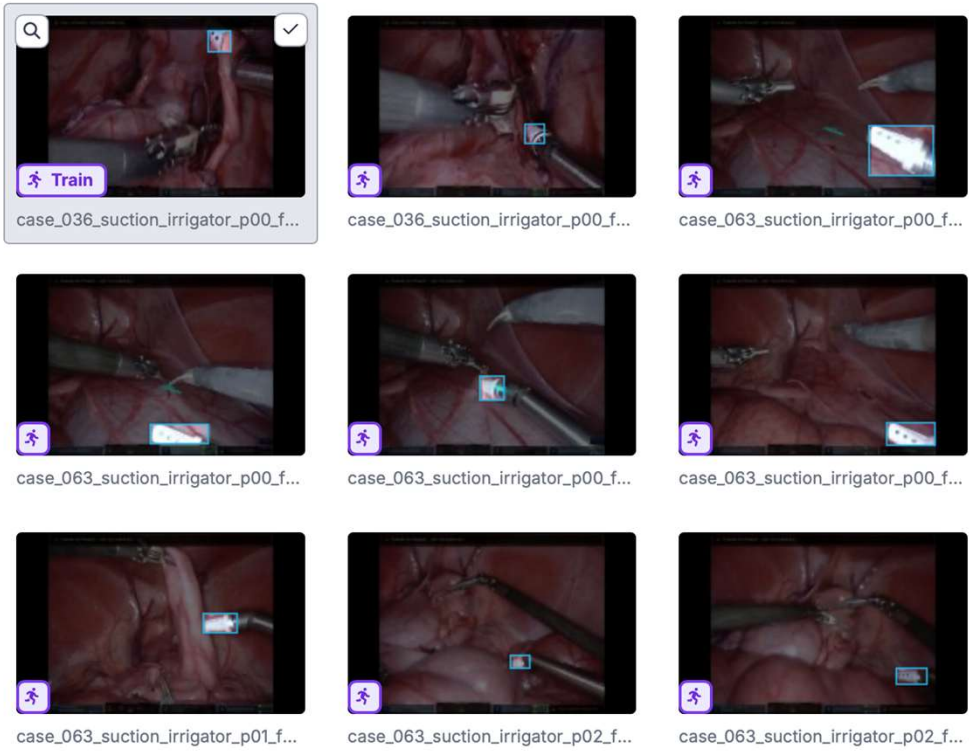
- SurgVU Val Set (Bounding Boxes)
  - “Of the twelve tools, 8 are contained” in the set
- 155 Video Session (No Bounding Boxes)
  - “tool labels can be considered noisy”



Manual Annotation



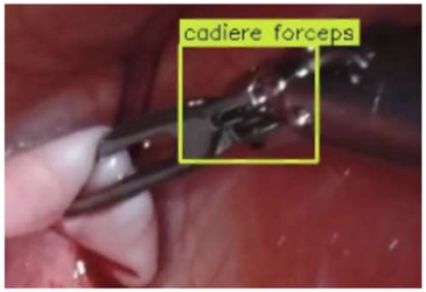
Annotation Examples Provided by Challenge



532 Frames Annotated Using Roboflow

Pseudo-Labels

Predictions



RF-DETR Teacher 0.5082 mAP@50:95

Per-Class Confidence Floors: 0.15-0.35 by Class

Session Cap: Max 15% per session

Class Quoata: 2800 boxes/class

Curated Pseudo-Labels

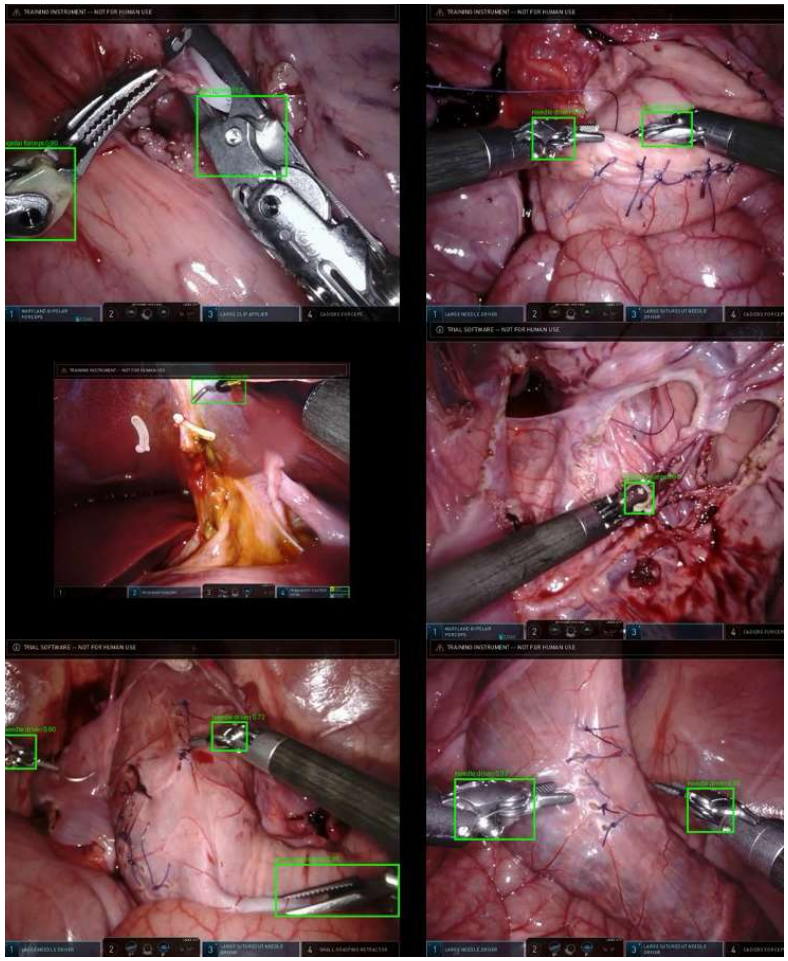
26,918 frames added to training set

Curated Hybrid Training Set

- Train Set: Balanced Integration:
    - **Human Annotated Labels:** 4,123 annotated frames
    - **Curated Pseudo-Labels:** 26,918 model-annotated frames
- Validation Set:
    - **Human-Annotated Only:** 1,588 annotated frames

Final Results

Model	mAP@50:95	Phase / dataset
Yolov8l	0.272	SurgVU Validation Set Spit
RF-DETR-L	0.352	SurgVU Validation Set Spit
RF-DETR-L (teacher)	0.5082	13-class Manual Annotations + SurgVU Validation Set Spit
RF-DETR-L	0.7979	Curated hybrid dataset
RF-DETR-L (best)	**0.8116**	**Curated hybrid dataset + 3:1 Ratio + Weight Decay**
D-FINE-L	0.7805	Curated hybrid dataset



Predictions from the final model (0.8116 mAP) across held-out validation frames

Thank You!