SURGVU 2026 · WEEK 7 PRESENTATION

Surgical Visual Understanding 2026 Category 1 Challenge

DeAndre Bailey

Project Recap

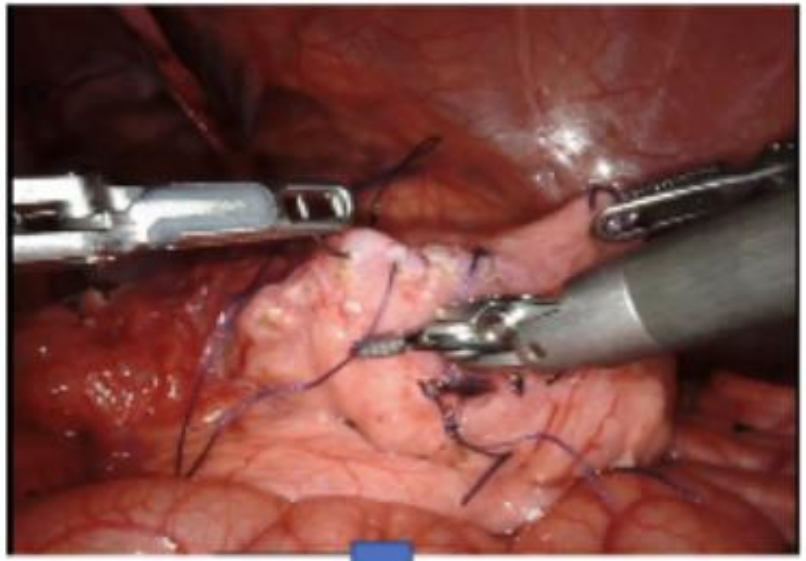
- Surgical Visual Understanding 2026 Category 1 Challenge
- Classify and localize surgical instruments using bounding boxes
- Develop **semi-supervised or weakly supervised models**



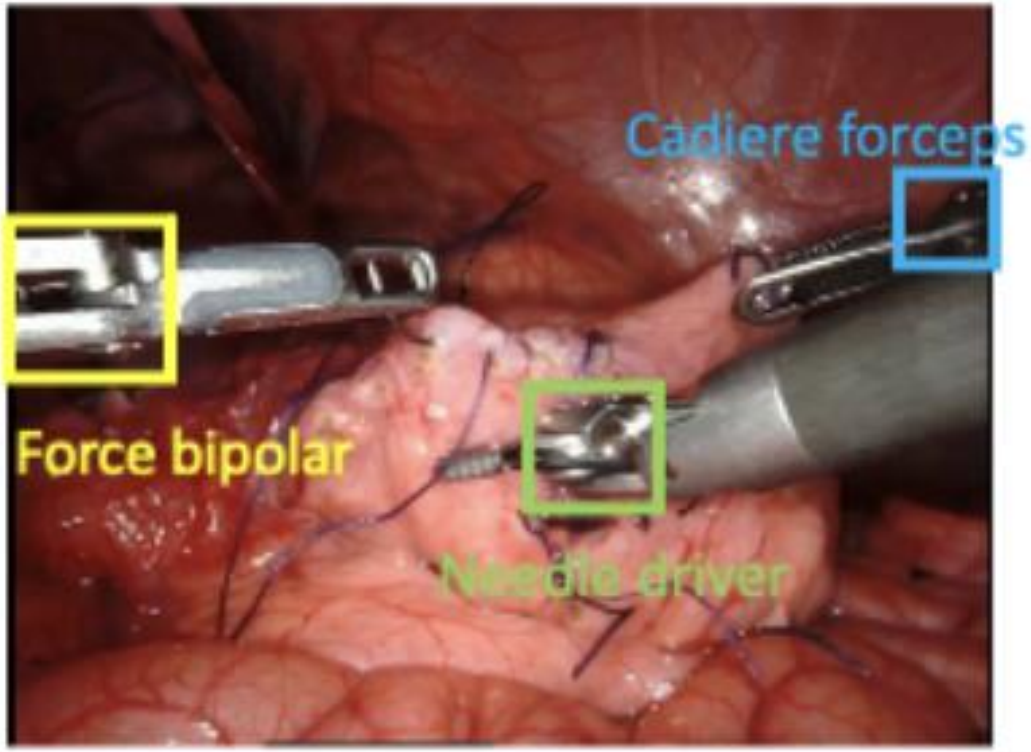
Project Recap

Category 1
(weak/noisy labels)

Input image

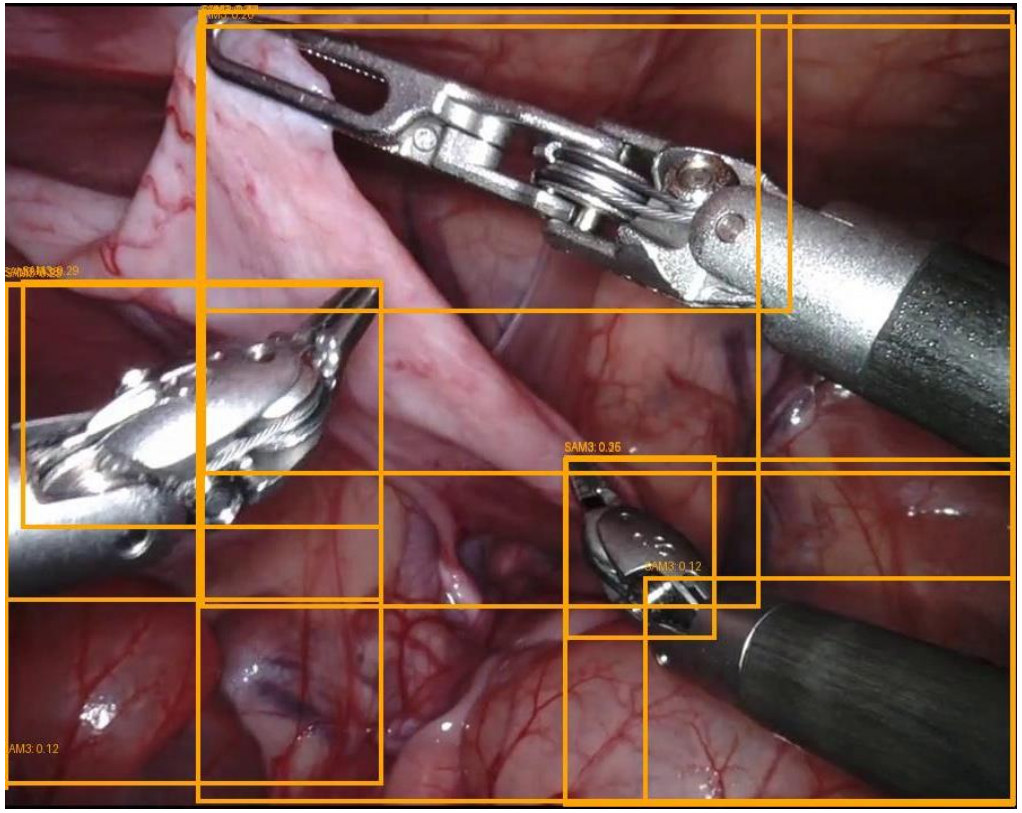


Classify and localize tools



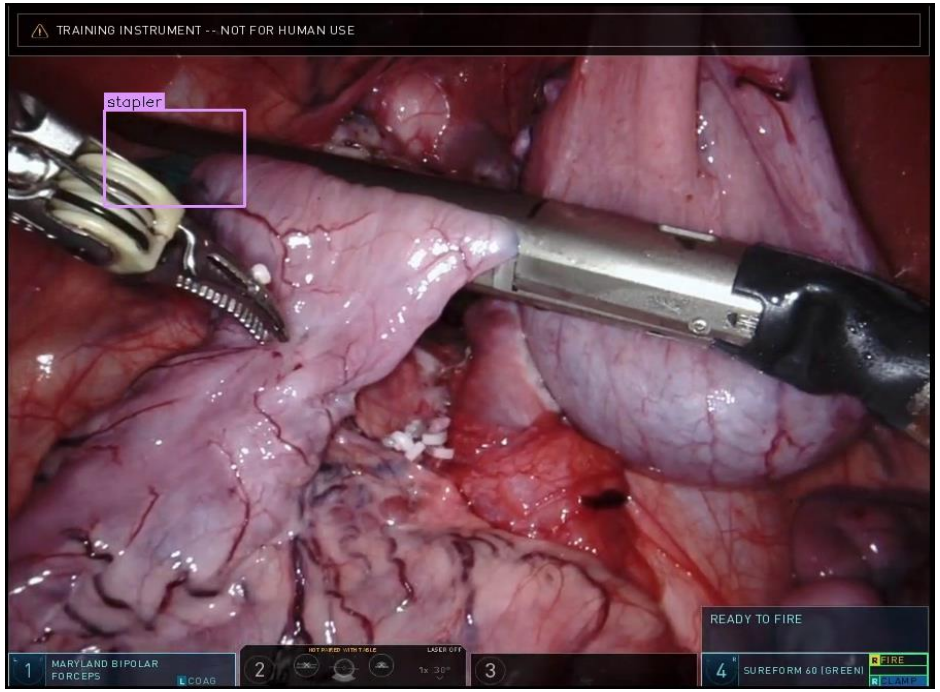
Last Week

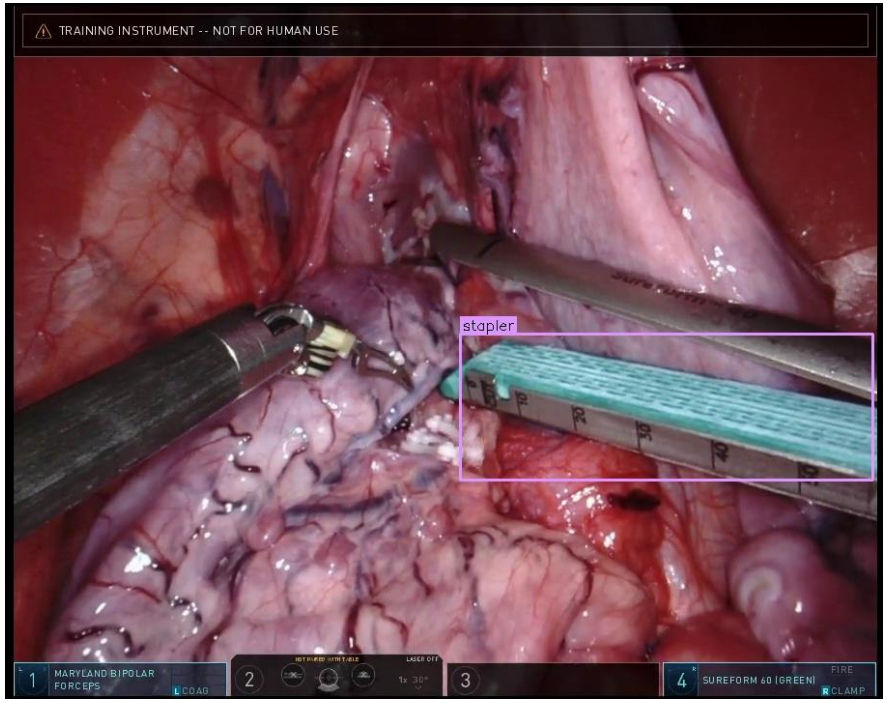
- Testing Annotations Methods (SAM3, Owlv2, Roboflow)
- Manually Annotate missing Classes
- Trained RF-DETR On New Annotations



Predictions (Round 1)

- Had the Model make predictions over case_001 (video session)
- Errors with Stapler and Tip-Up Grasper annotations



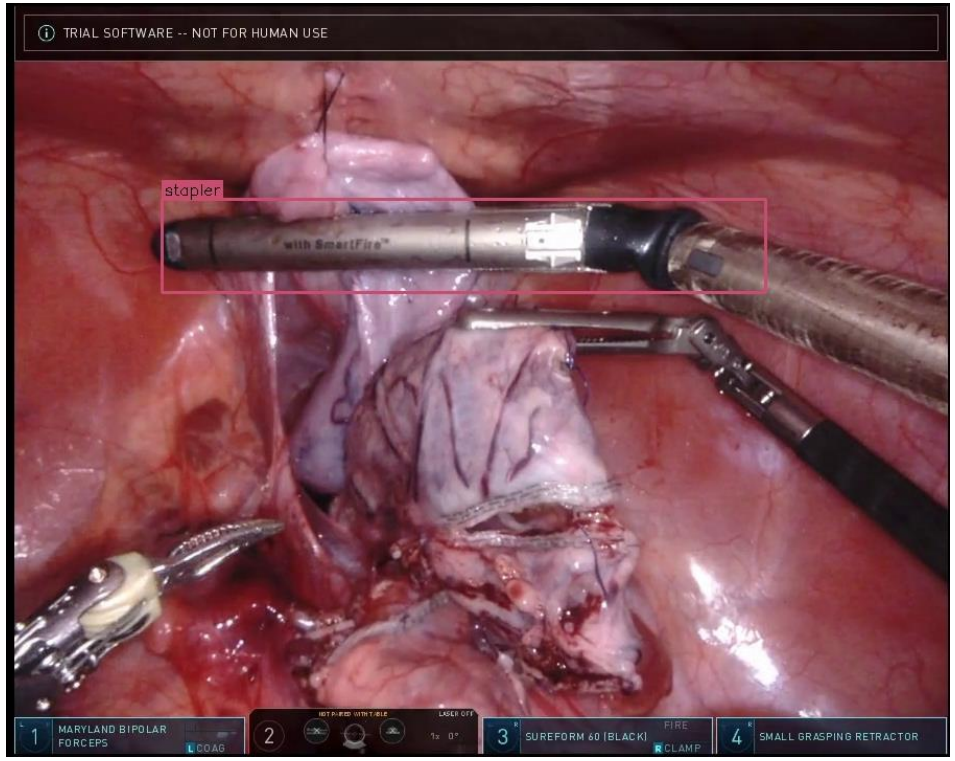


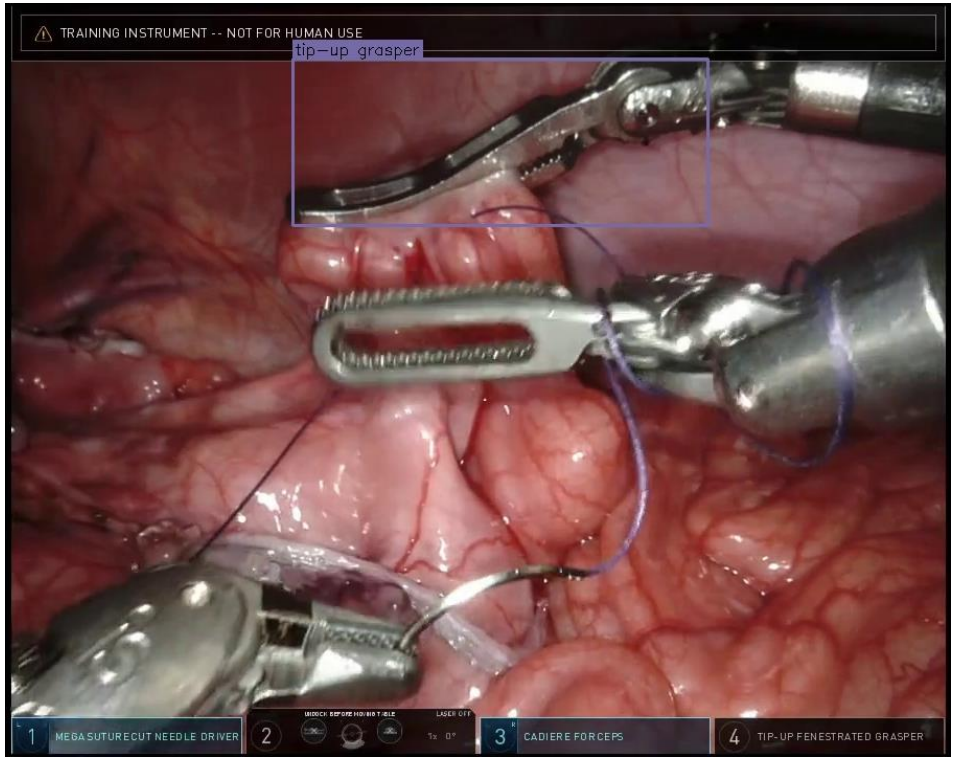
Correcting & Re-Annotating

- Manually re-annotated **72 mislabeled frames**
  - Fixed via Roboflow, exported and Merged into training Dataset
- Retrain RF-DETR on corrected data

Predictions (Round 2)

- Model make Prediction over case_001 again
- Produced Better Annotations for Stapler and Tip Up





RF-DETR V3 Metrics

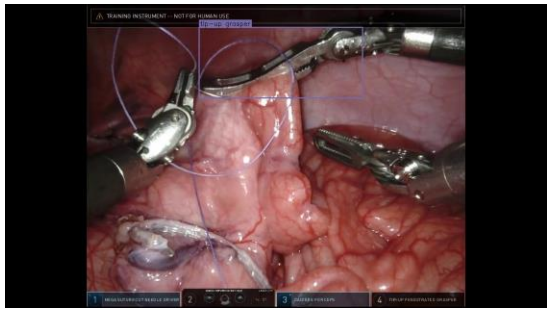
Metric	Value
mAP@50:95	**0.5082**
Base checkpoint	mAP@50:95 = **0.4549**
Improvement (from Base)	+0.0533 mAP@50:95

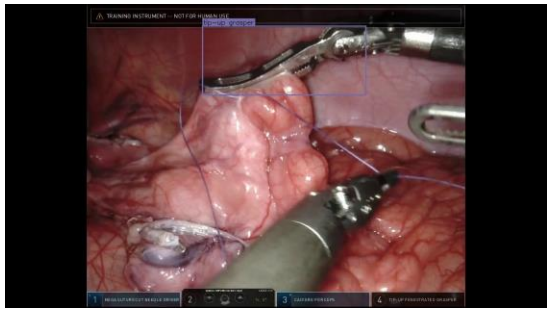
Class	AP@50:95	Precision	Recall
Cadiere forceps	**0.0267**	0.3730	0.1366
Clip applier	**0.1177**	0.0000	0.0000
Stapler	**0.2371**	0.5400	0.5400

Pseudo Labels

Predictions

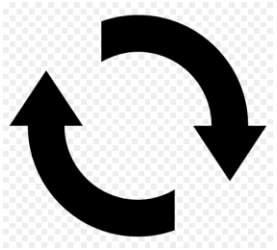
RF-DETR or SeqTrack





Train

Merge to Current Dataset



Retrain Final Model (RF-DETR)

Pseudo-Labeling

- Ran across all **155 sessions**
  - 1 frame per second is sampled
  - Detected class must be in list of installed tools
  - Per-class confidence threshold
    - Remove Annotations with low confidence
- Generated **539,560** pseudo-labels total

Per-Class Threshold

Class	Threshold	Reason
Cadiere forceps	**0.15**	Worst-performing class. Lowered from the 0.35 default specifically
Clip applier	**0.20**	”Struggling" class
Stapler	**0.20**	”Struggling" class
11 Other Classes	**0.35**	Standard/default bar — (0.4–0.9 range per your per-class tables)

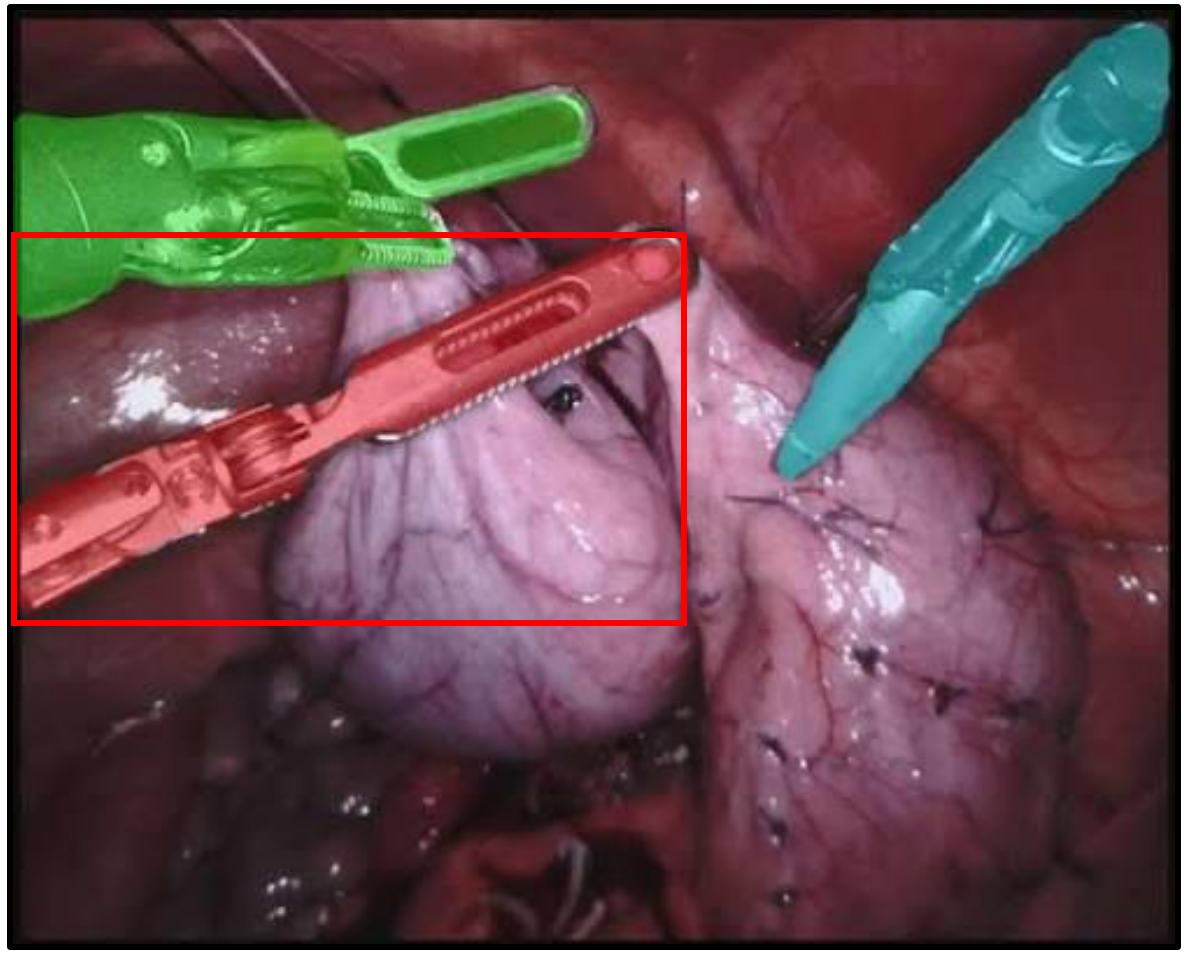
Pseudo-Labeling

- **539,560** pseudo-labels total is too many
- This Week:
  - How to reduce this number Further?

Tool	Pseudo-Labels
Monopolar curved scissors	175,530
Prograsp forceps	166,871
Needle driver	97,020
Cadiere forceps	42,817
Vessel sealer	31,132
Bipolar forceps	12,354
Permanent cautery hook/spatula	5,800
Force bipolar	2,573
Stapler	2,218
Tip-up fenestrated grasper	1,577
Grasping retractor	866
Suction irrigator	537
Clip applier	391
Bipolar dissector	0

Exploring Label Propagation

- Moving Bounding Box forward in frames
- SAM 3.1 (Other Segmentation Models)
- Create bounding box around the entire tool
- Need specific Box around Clevis



Exploring Label Propagation Pt. 2

Less is More: Token Context-aware Learning for Object Tracking

Chenlong Xu[1], Bineng Zhong[1]*, Qihua Liang[1], Yaozong Zheng[1]
Guorong Li[2], Shuxiang Song[1]
[1]Key Laboratory of Education Blockchain and Intelligent Technology, Ministry of Education,
Guangxi Normal University, Guilin 541004, China
[2]Key Laboratory of Big Data Mining and Knowledge Management, University of Chinese Academy of Sciences
xuchenlong@stu.gxnu.edu.cn, bnzhong@gxnu.edu.cn, qhliang@gxnu.edu.cn, yaozongzheng@stu.gxnu.edu.cn,
liguorong@ucas.edu.cn, songshuxiang@mailbox.gxnu.edu.cn

- Less is More: Token Context-aware Learning for Object Tracking
- **OSTrack (PUH-SVU):**
  - Stable position, but box shape drifted from the seed
- **SeqTrack:**
  - Reports a **confidence score per frame** — a real drift signal
  - Confidence trended down over the propagation window

Exploring Label Propagation Pt. 3

Less is More: Token Context-aware Learning for Object Tracking

Chenlong Xu[1], Bineng Zhong[1]*, Qihua Liang[1], Yaozong Zheng[1]
Guorong Li[2], Shuxiang Song[1]
[1]Key Laboratory of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, Guilin 541004, China
[2]Key Laboratory of Big Data Mining and Knowledge Management, University of Chinese Academy of Sciences
xuchenlong@stu.gxnu.edu.cn, bnzhong@gxnu.edu.cn, qhliang@gxnu.edu.cn, yaozongzheng@stu.gxnu.edu.cn, liguorong@ucas.edu.cn, songshuxiang@mailbox.gxnu.edu.cn

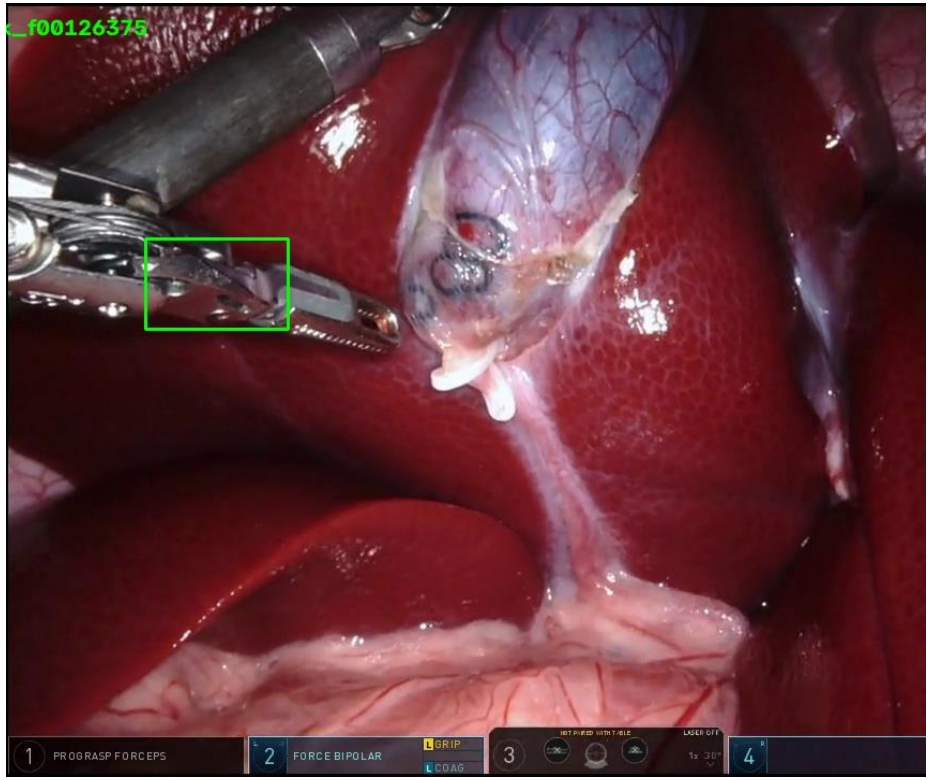
- LMTrack looked stronger on paper, but ships no pretrained weights
  - If time permits worth taking a look at training
- Chose SeqTrack for its built-in confidence signal and pretrained weights

Method	GOT-10k AO	LaSOT AUC	TrackingNet AUC
OSTrack384	73.7	71.1	83.9
SeqTrack384	74.5	71.5	83.9
LMTrack384	**80.1**	**73.2**	**85.7**

SeqTrack Testing

- Confidence Decreases overtime when Propagating

```
case case_070 cls 7: filling 84.0s gap
  frame 950340: confidence 0.021 below
case case_070 cls 7: filling 58.0s gap
  frame 956220: confidence 0.050 below
case case_070 cls 7: filling 90.0s gap
  frame 961080: confidence 0.073 below
```



Next Steps

- Complete the SeqTrack Testing
- Merge **pseudo-labels** into one training set
- Final retrain and evaluation
- Revisit Classes with weak AP Scores
  - Data Augmentation and Datasets
  - Manual Annotations

Questions?