SURGVU 2026 · WEEK 9 PRESENTATION

# Surgical Visual Understanding 2026 Category 1 Challenge

*DeAndre Bailey*

# Project Recap

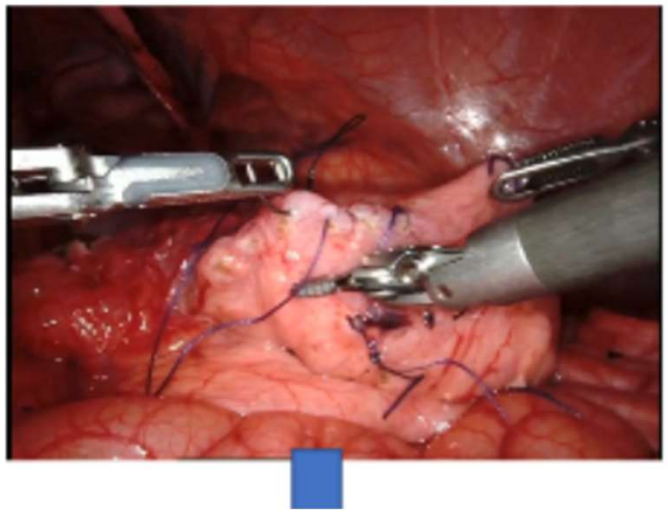
- Surgical Visual Understanding 2026 Category 1 Challenge
- Classify and localize surgical instruments using bounding boxes
- Develop **semi-supervised or weakly supervised models**



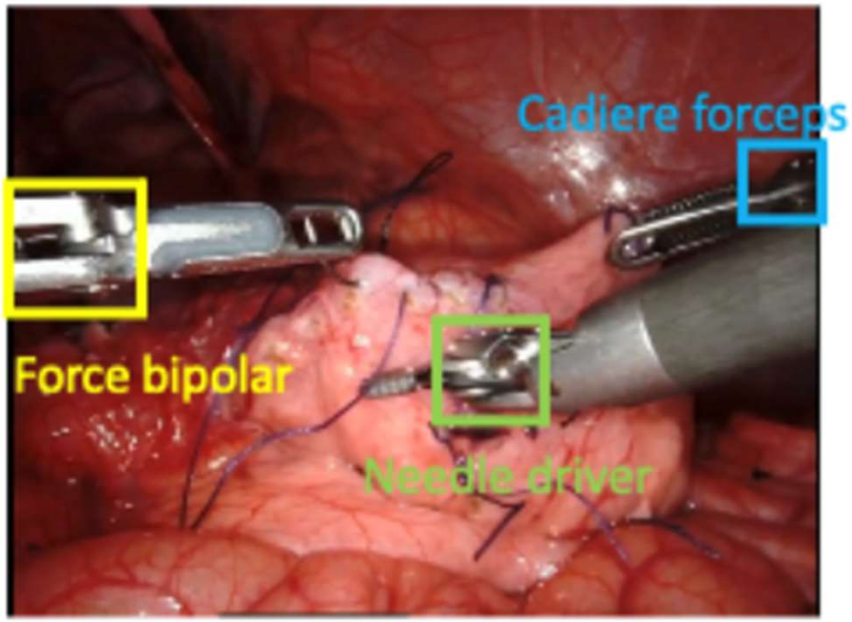
# Project Recap

**Category 1**
**(weak/noisy labels)**

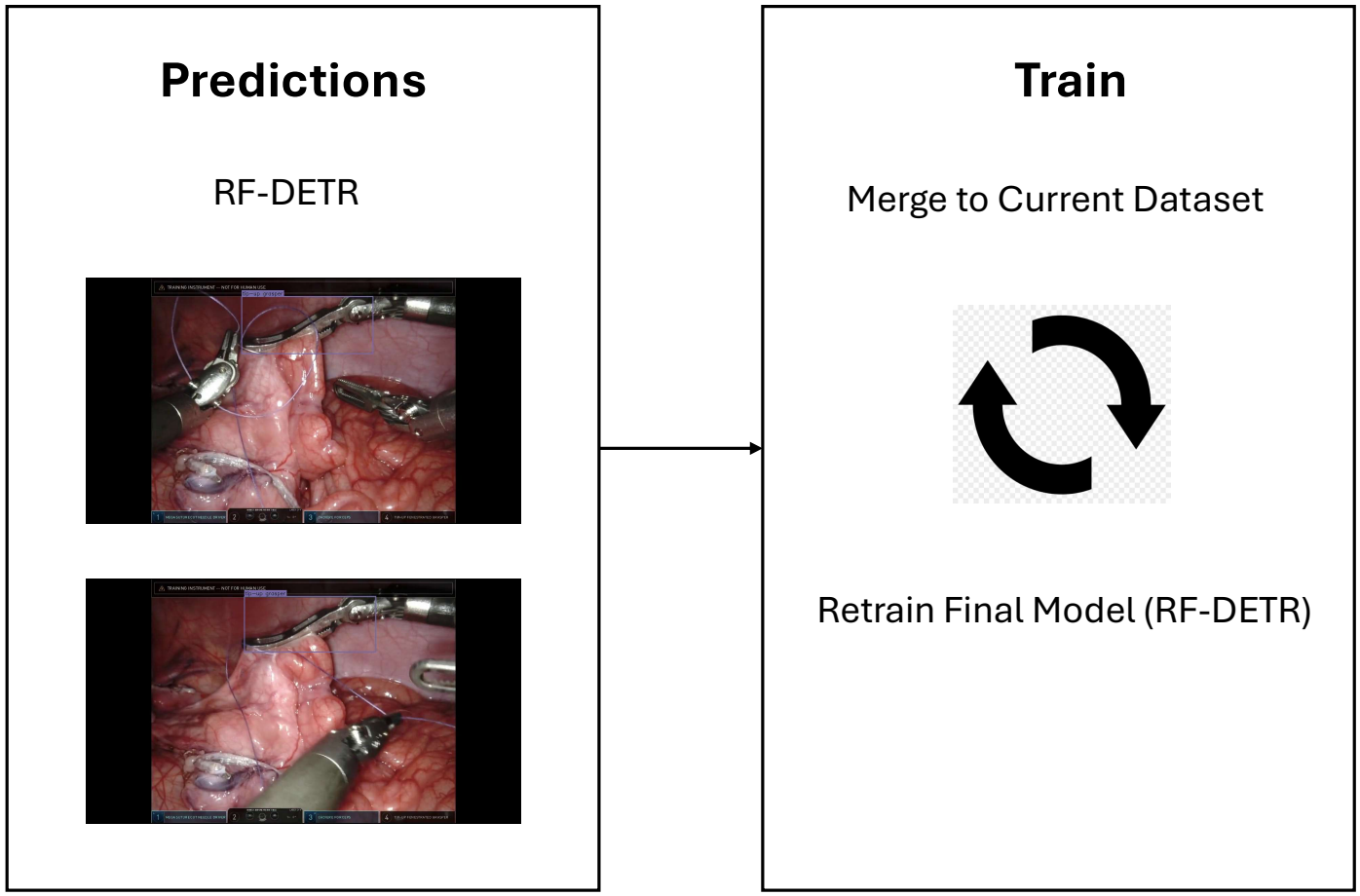
Input image



Classify and localize tools

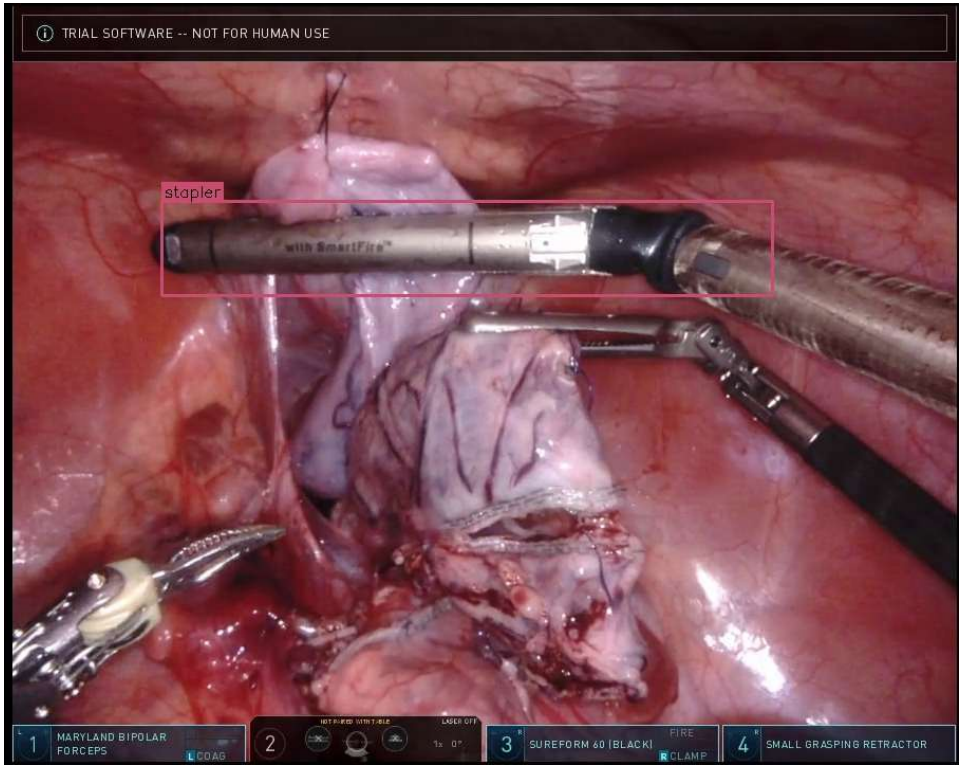


# Pseudo-Labels



# Last Week

- **Created Hybrid Training Set**
- RF-DETR V4 mAP @ 50:95 at **0.7979**
- **Prograsp forceps/tip-up grasper** flagged as weak with
  - No diagnosis last week



# Weak Classes

- Tip up fenestrated grasper and Prograsp Forceps had the Lowest AP Scores (RF-DETR V4)

| Class | AP 50:95 |
|---|---|
| **tip up fenestrated grasper** | 0.5764 |
| **prograsp forceps** | 0.5674 |

- Is it *not finding* the tool or *mislabeling it*?

# Weak Classes

- Localization and Classification for Prograsp Forceps and Tip Up Fenestrated Grasper

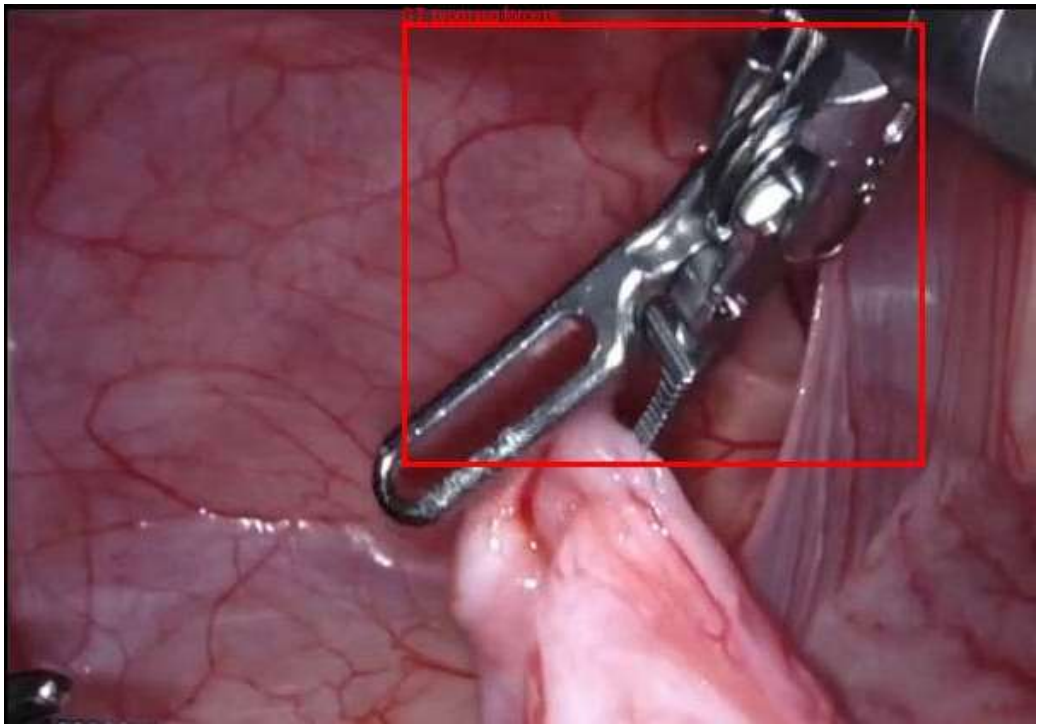
| Class | Localization (IoU Threshold 0.5) | Classification Accuracy |
|---|---|---|
| Prograsp forceps | **0.854** | 0.976 |
| Tip-up fenestrated grasper | **0.778** | 1.000 |

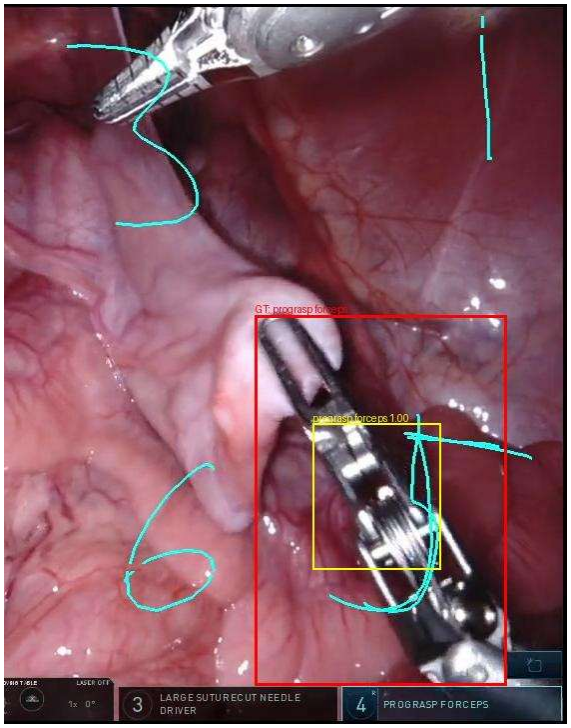
- Localization Error

# Localization and Classification (All Classes)

| Class | Localization (IoU Threshold 0.5) | Classification Accuracy |
|---|---|---|
| Tip-up fenestrated grasper | 0.778 | 1.000 |
| Prograsp forceps | 0.854 | 0.976 |
| Stapler | 0.929 | 1.000 |
| Permanent cautery hook/spatula | 0.929 | 1.000 |
| Clip applier | 0.968 | 1.000 |
| Monopolar curved scissor | 0.980 | 1.000 |
| Needle driver | 0.988 | 1.000 |
| Cadiere forceps | 0.990 | 0.998 |
| Bipolar forceps | 0.994 | 1.000 |
| Force bipolar | 0.996 | 1.000 |
| Grasping retractor | 1.000 | 1.000 |
| Vessel sealer | 1.000 | 1.000 |
| Suction irrigator | 1.000 | 1.000 |

# Testing Hypotheses





# Ablation

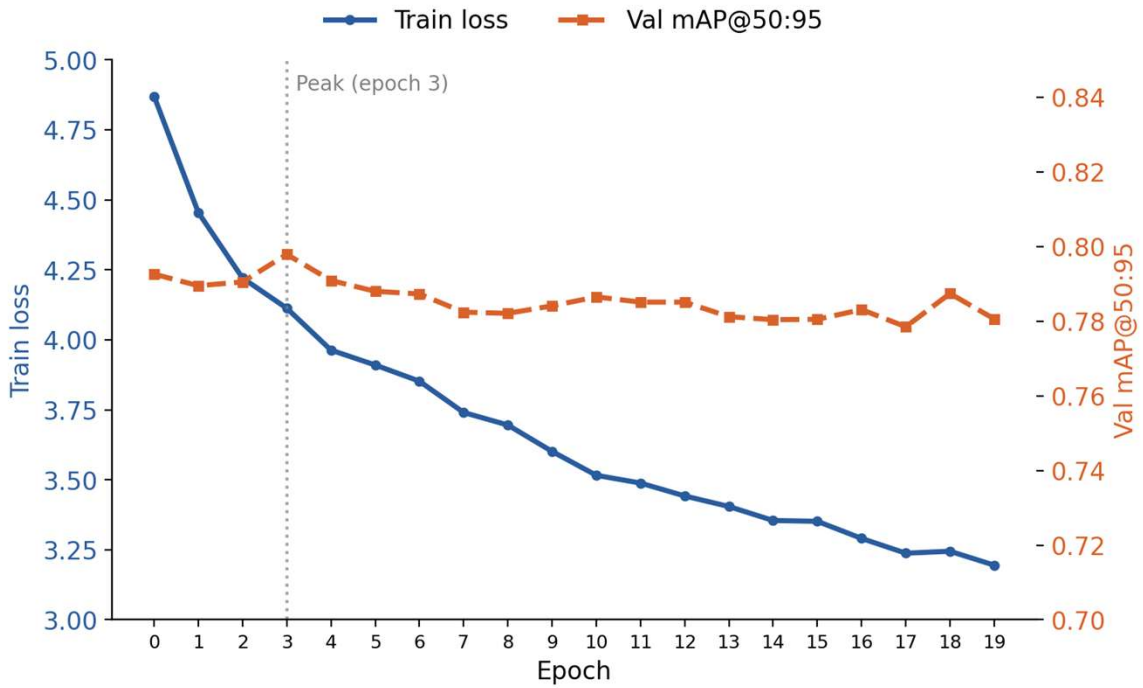
| Class | Pseudo-Labels 15% Session Cap | Pseudo-Labels No Cap |
|---|---|---|
| Tip-up fenestrated grasper | 638 | 1,747 |
| Grasping retractor | 718 | 1,374 |
| Suction irrigator | 281 | 665 |
| Clip applier | 549 | 549 |

| Class | Discounted Floor | Flat 0.35 Floor | Change |
|---|---|---|---|
| Clip applier | 549 (0.20) | 211 | **−338 (−62%)** |
| Grasping retractor | 718 (0.20) | 378 | **−340 (−47%)** |

- Next: Currently Training RF-DETR on these Ablation Datasets

# Overfitting Issue

- Best result at epoch 3
- No improvement over the next 17 epochs
- Train loss fell ~35%
- Val mAP stayed flat
- Sign of Overfitting



# Overfitting Issue

- Two-part fix, applied together:
  - Reduced pseudo:ground-truth ratio (~6.5:1 → ~3:1)
  - Added weight decay (1e-4 → 5e-4)

| Metric | mAP@50:95 |
|---|---|
| RF-DETR V4 (previous best) | 0.7979 |
| **New Best (Ratio + Weight Decay)** | **0.8116** |
| Improvement | **+0.0137** |

- The best checkpoint landed very early (epoch 1)

# D-FINE Metrics

- D-FINE is a real-time object detector
- D-FINE and RF-DETR both trained on Hybrid Training Set
  - Without **Ratio + Weight Decay**

| Model | Starting point | mAP@50:95 |
|---|---|---|
| D-FINE | Generic COCO pretraining | **0.7805** |
| RF-DETR (generic-control) | Generic pretraining (matched, same starting line) | **0.7254** |

# Next Steps

- Complete Report and Poster
- Test Decoupled Architecture
- Test D-FINE on

# Questions?