SURGVU 2026 · WEEK 8 PRESENTATION

# Surgical Visual Understanding 2026 Category 1 Challenge

*DeAndre Bailey*

# Project Recap

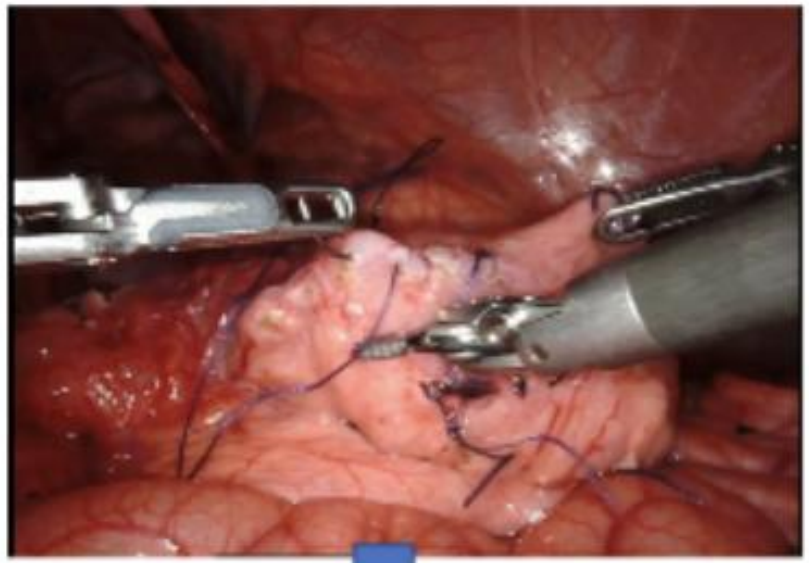
- Surgical Visual Understanding 2026 Category 1 Challenge
- Classify and localize surgical instruments using bounding boxes
- Develop **semi-supervised or weakly supervised models**



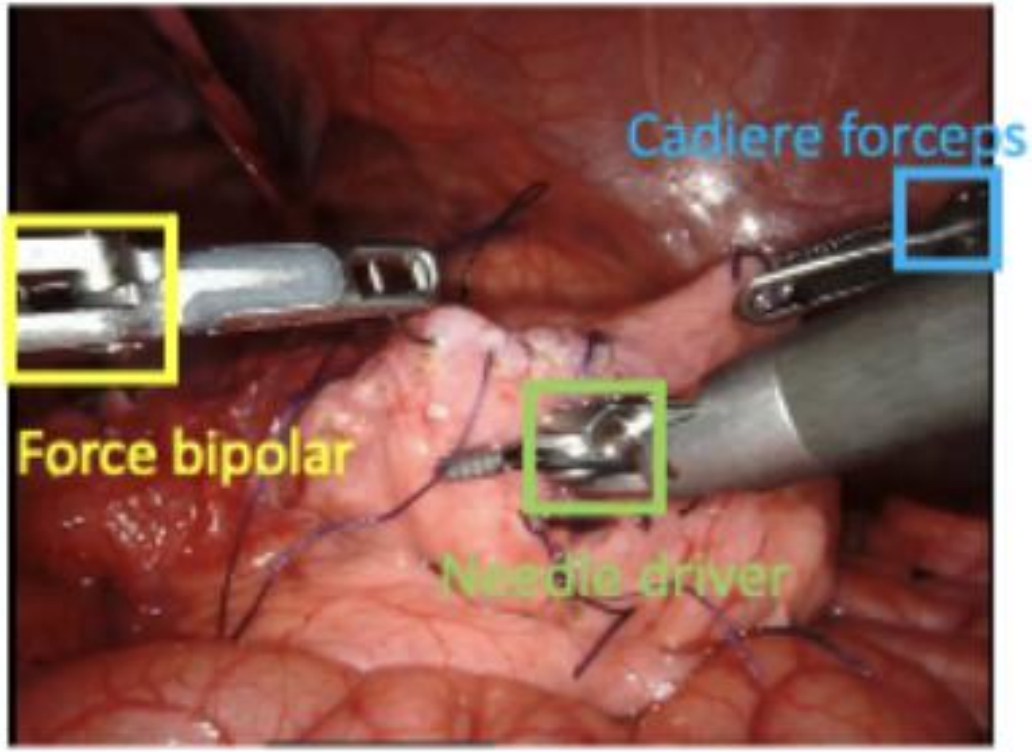
# Project Recap

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

Input image

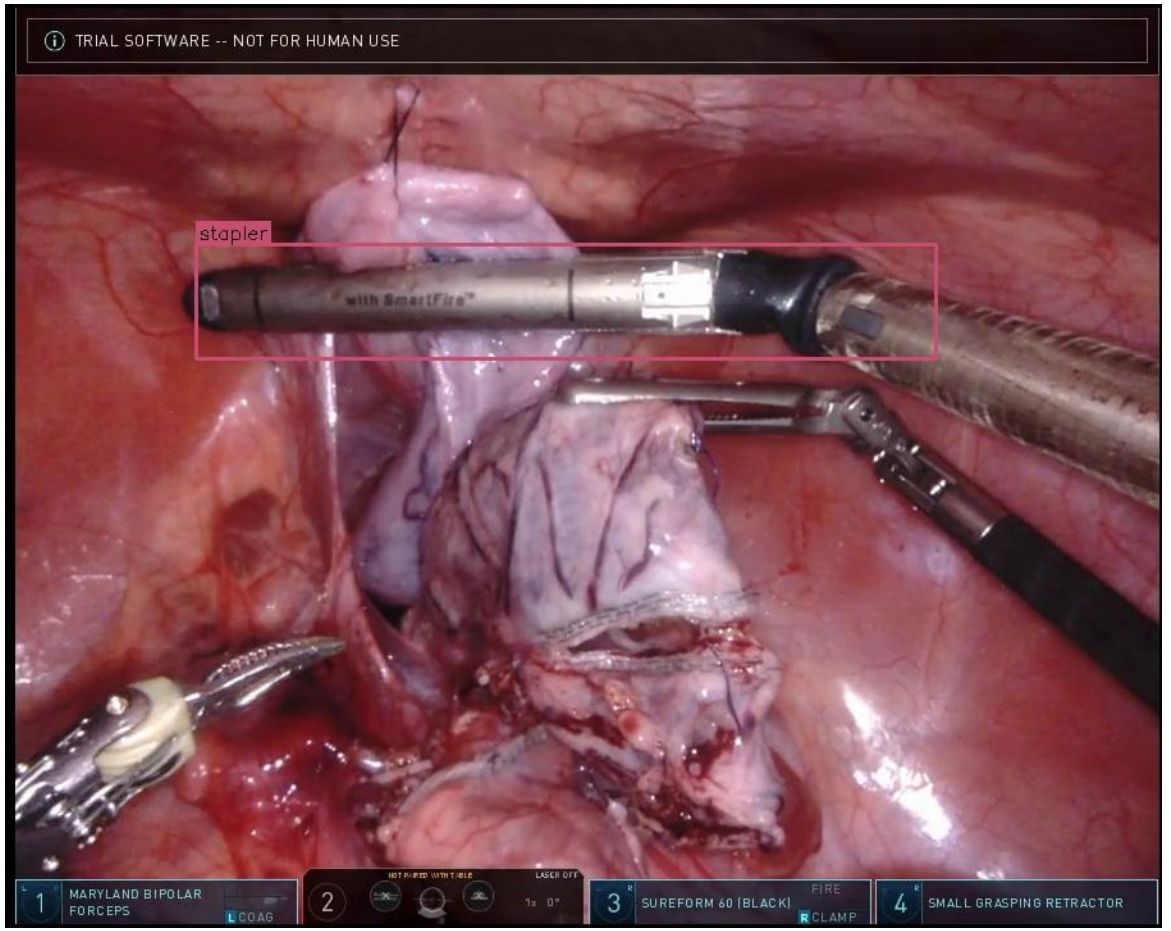


Classify and localize tools



# Last Week

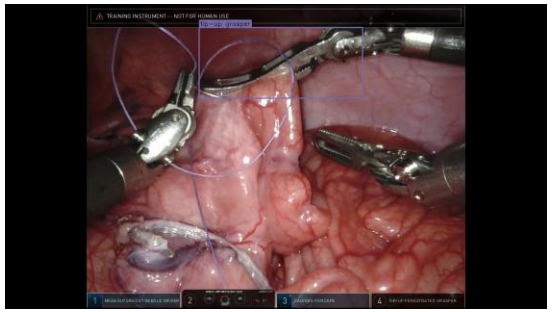
- Trained RF-DETR V3 as the pseudo-label source model
- 539,560 candidate pseudo-labels
- Exploring Label Propagation

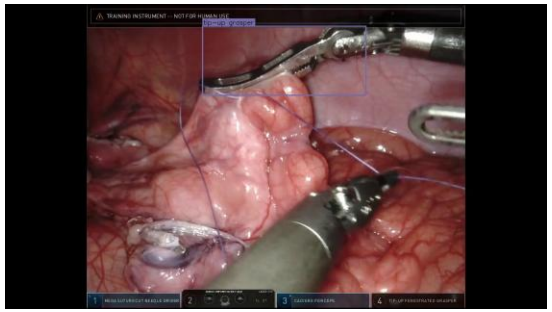


# Pseudo-Labels

## Predictions

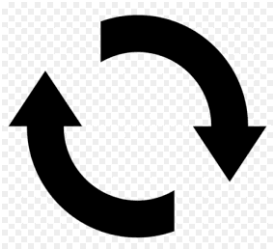
RF-DETR or SeqTrack





## Train

Merge to Current Dataset



Retrain Final Model (RF-DETR)

# Issue with Previous Pseudo-labels

- **539,560** Pseudo-labels from RF-DETR Predictions

- How do we get the *best?*:
  - Eliminate Noise
  - Fix Class Imbalance
  - Prevent Redundant Frames (Diversity)

# Eliminating Noise

- Avoid reinforcing the model's own mistakes
- Per-class confidence thresholds
- Lower Threshold for Weaker classes

| Class | AP@50:95 |
|---|---|
| Cadiere forceps | **0.0267** |
| Clip applier | 0.1177 |
| Stapler | 0.2371 |

| Class | Threshold |
|---|---|
| Cadiere forceps | **0.15** |
| Clip applier | **0.20** |
| Stapler | **0.20** |
| 11 Other Classes | 0.35 |

# Prevent Redundant Frames (Diversity)

- Avoid overfitting to a handful of sessions
- Random session order during selection
- Cap: no session contributes more than 15% of a class's boxes



# Fix Class Imbalance

- Capped-Proportional Quotas

- 2800 boxes Flat Quota per-class
  - Any additional high confidence boxes in frame are kept

- Rare classes had too few candidates to hit a flat quota

  - Tip-up fenestrated grasper (**638 boxes from 3 sessions**)
  - Suction irrigator (**281 boxes from 5 sessions**)

# Curated Hybrid Training Set

- Train Set: Balanced Integration
  - **Human GT Baseline:** 4,123 annotated frames
  - **Curated Pseudo-Labels:** 26,918 model-annotated frames (Diversity-sampled & quality-filtered).

- **Validation Set: Human-Annotated Only**
  - 1,588 frames
  - Strictly human-verified to ensure reliability

# Training Set Class Split

- **31,041** frames labeled
- **42,605** Train Boxes
- **3,459** Val Boxes

| Class | Train Boxes | Val Boxes |
|---|---|---|
| monopolar curved scissors | 8,851 | 760 |
| needle driver | 6,188 | 1,033 |
| cadiere forceps | 5,330 | 591 |
| bipolar forceps | 4,243 | 511 |
| force bipolar | 3,351 | 257 |
| prograsp forceps | 3,292 | 48 |
| vessel sealer | 3,006 | 33 |
| stapler | 3,004 | 28 |
| permanent cautery hook/spatula | 2,847 | 14 |
| grasping retractor | 859 | 70 |
| tip-up fenestrated grasper | 684 | 9 |
| clip applier | 654 | 94 |
| suction irrigator | 296 | 11 |
| **Total** | **42,605** | **3,459** |

# RF-DETR V4 Metrics

- Performed Better than RF-DETR **+0.2897**
- Potential Issue of overfitting
- Best result at epoch 3; no improvement over the next 17 epochs

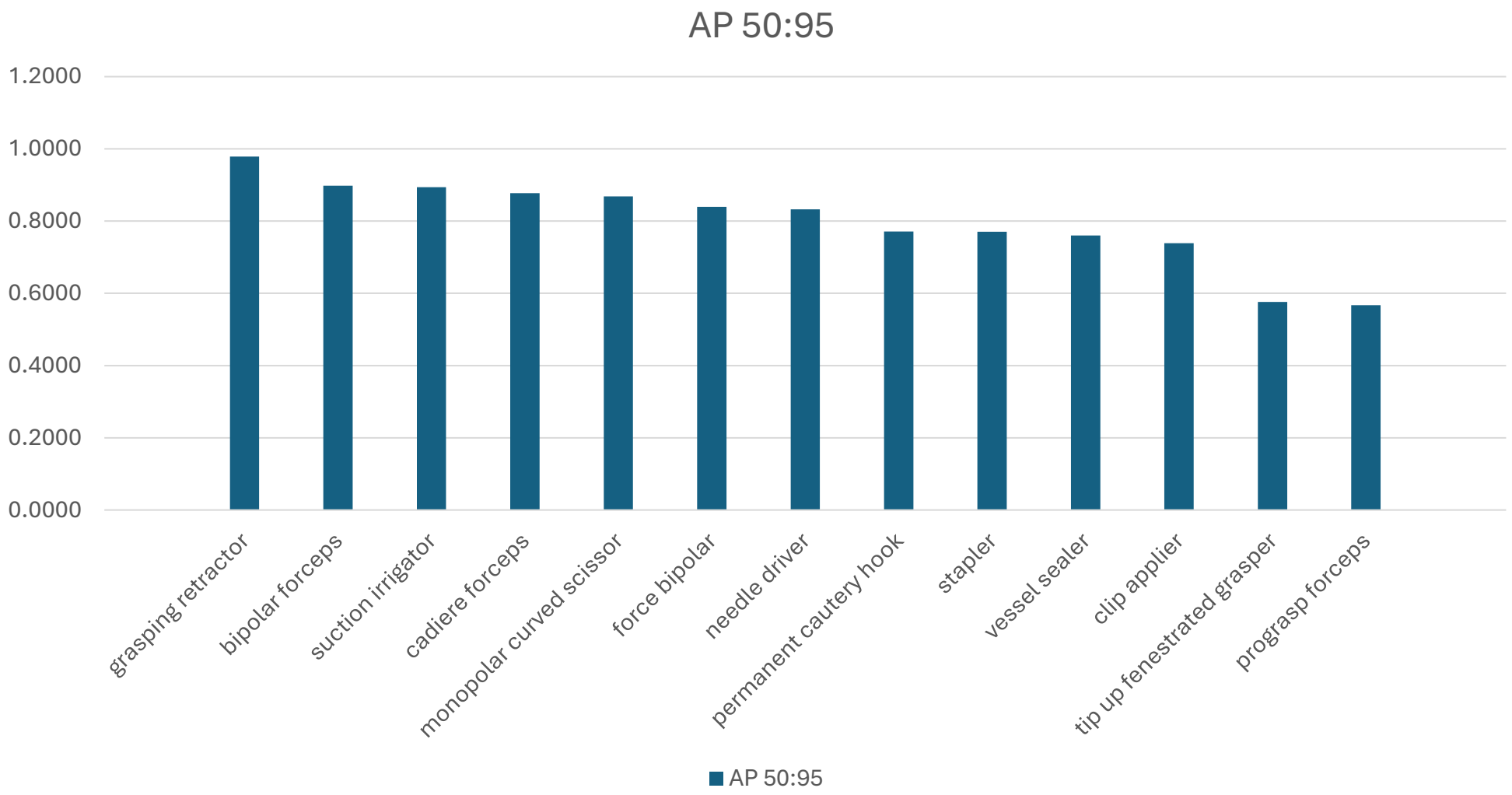
| Metric | mAP@50:95 |
|---|---|
| **RF-DETR V4** | **0.7979** |
| RF-DETR V3 | 0.5082 |
| Improvement (from RF-DETR V3) | **+0.2897** mAP@50:95 |

# RF-DETR V4 Class Metrics

| Class | AP 50:95 |
|---|---|
| **grasping retractor** | 0.9784 |
| **cadiere forceps** | 0.8772 |
| **bipolar forceps** | 0.8983 |
| **suction irrigator** | 0.8938 |
| **force bipolar** | 0.8393 |
| **stapler** | 0.7702 |
| **permanent cautery hook** | 0.7712 |

| Class | AP 50:95 |
|---|---|
| **monopolar curved scissor** | 0.8685 |
| **vessel sealer** | 0.7602 |
| **needle driver** | 0.8324 |
| **clip applier** | 0.7391 |
| **tip up fenestrated grasper** | 0.5764 |
| **prograsp forceps** | 0.5674 |

# RF-DETR V4 Class Metrics



# Next Steps

- Test decoupled classifier comparison
- Investigate Prograsp forceps and tip-up fenestrated grasper underperformance
- Complete Report and Poster

# Questions?