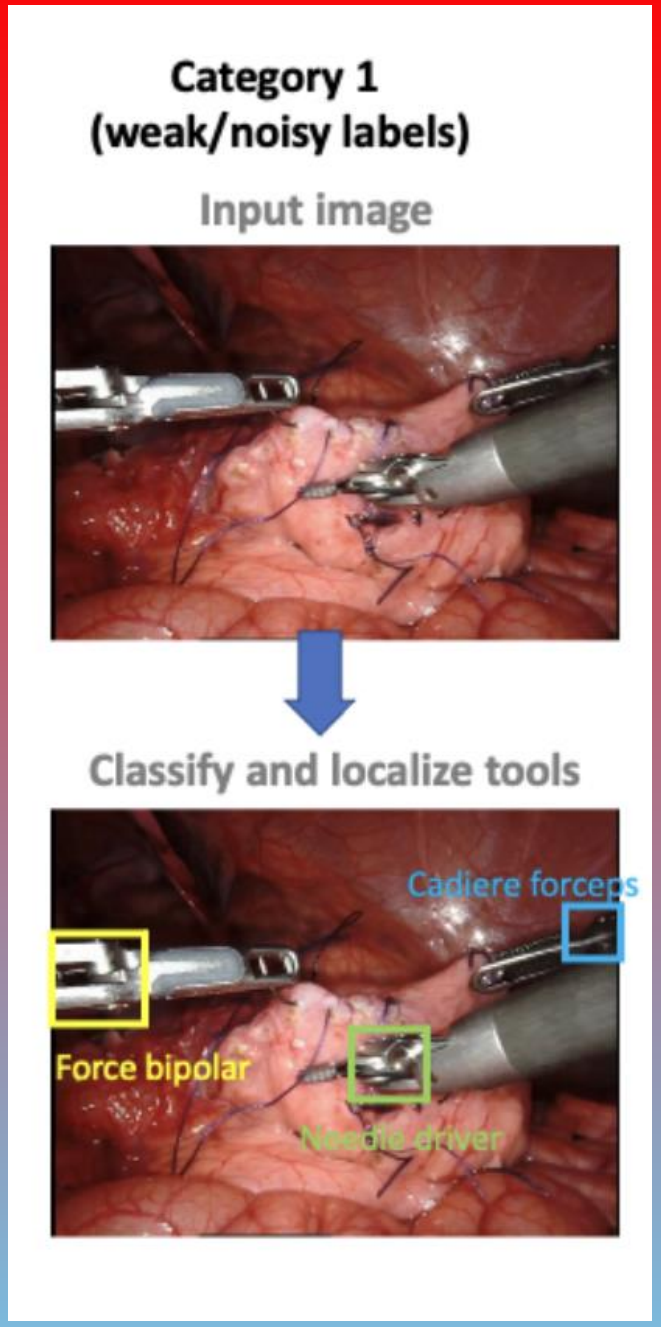
SURGVU 2026 · WEEKLY PROGRESS REVIEW

Week 5 Presentation

DeAndre Bailey

CRCV REU · SurgVU Challenge 2026



Project Recap

- **SurgVU Challenge**: Category 1 Classification and Localization
- 12 official tool classes that must be detected
- mAP@50 is the primary leaderboard score
- Last week: Missing Ground Truths for 4 Tools (little ground truths for others)
- This week: Extracting Candidates and 4 Model Evaluation

Frame Method Overview

Need to choose the best frames for these 4-5 missing classes

Take Candidates from Each Session

tools.csv provide where Tools are Present

Pool Candidates

Merged across all 155 sessions per class

K-Center-Greedy

Diversity selection via CNN features

Final Set

Top-N diverse frames for annotation

K-Center-Greedy (Sener & Savarese 2018)

❌ Random	❌ K-Means	✅ K-Center-Greedy
Redundant consecutive frames	Misses rare appearances	Guarantees coverage
No diversity guarantee	Biased toward dense regions	Budget-controlled

$$\min_{\mathbf{s}^1:|\mathbf{s}^1|\leq b} \max_{i} \min_{j\in\mathbf{s}^1\cup\mathbf{s}^0} \Delta(\mathbf{x}_i, \mathbf{x}_j)$$

- $\Delta(\mathbf{x}_{i'}\,\mathbf{x}_j)$= distance between frame i and frame j in CNN feature space
- Δ is Euclidean distance in YOLOv8l feature space (512-dim), $\mathbf{s}^0 = \emptyset$, budget b= 100–200 frames per class.

Filtering out Frames with few or no Tools

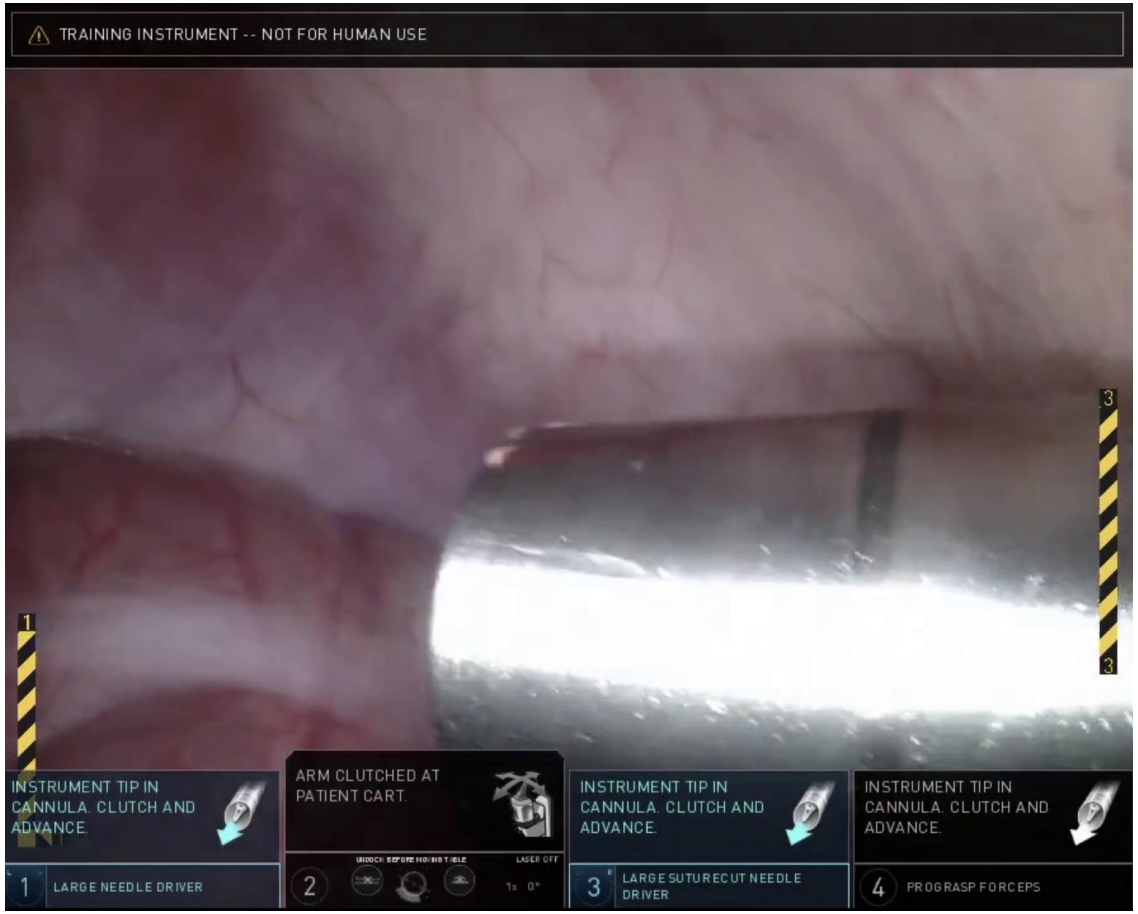
- 20-80% windowing tested
- Filtering out Frames with no tools
  - **Yolov8l Model hasn't seen some of these tools**

Confidence Threshold	Keep Rate
0.10	57.8%
0.25	50.9%
0.40	48.1%
0.60	42.1%

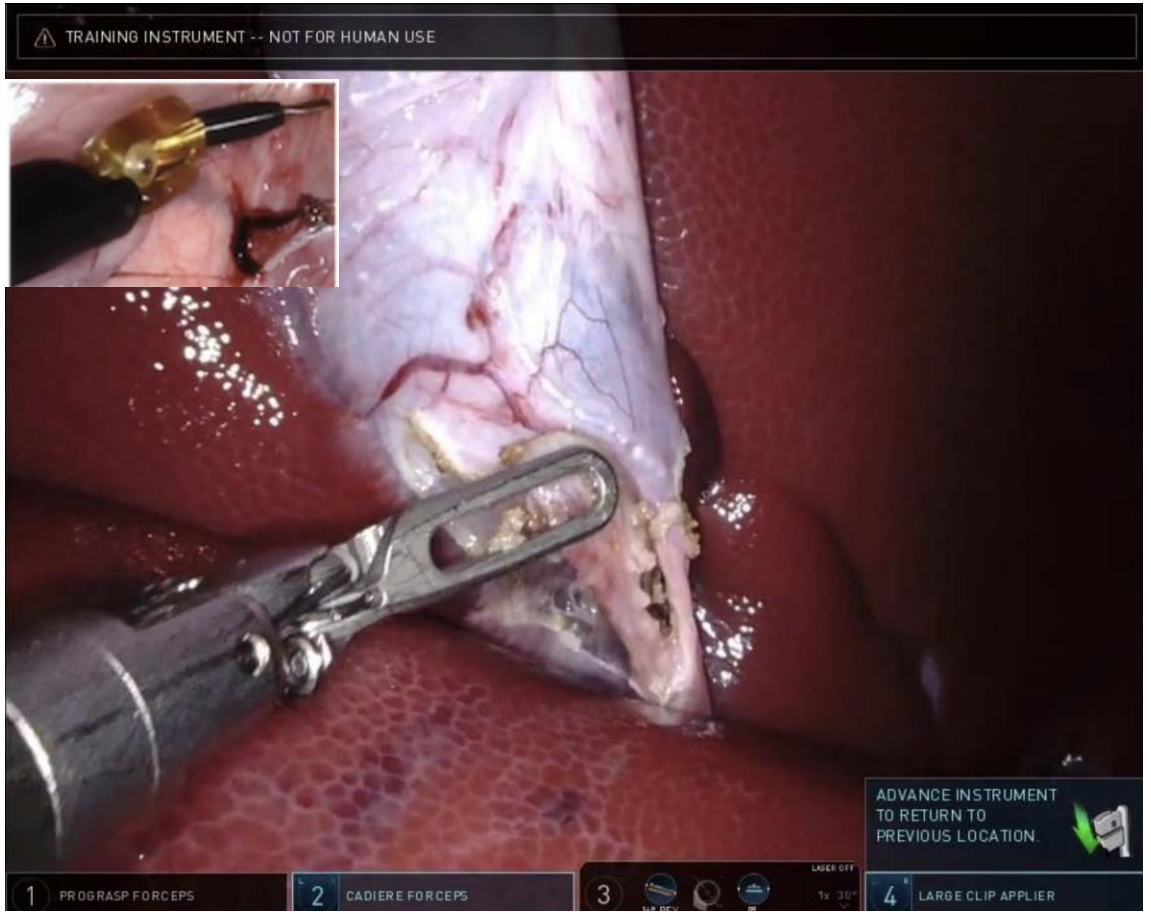
Conf=0.10, ~42% Frames Discarded
Using Single Yolov8l Single Tool Detector

- Automatic sampling + manual review

Bad Frames Extracted



Prograsp Forceps — case_055 (Obscured)



Permanent Cautery Hook — case_003 (Different Tool)

Annotation Progress

- 800 frames extracted across 5 missing classes
- Candidates pooled from 14–20 sessions per class
- K-Center-Greedy reduced thousands of candidates to 100–200 per class

Tool	Frames Extracted
Stapler	200 frames
Tip-up fenestrated grasper	200 frames
Permanent cautery hook	150 frames
Prograsp forceps	150 frames
Vessel sealer	100 frames
Total	**800 frames**

Model Evaluation 2 sessions (sess001 + sess006)

Model	mAP@50	Precision	Recall	Params
RT-DETR	**0.576**	**0.708**	**0.557**	32.0M
YOLOv11x	0.568	0.666	0.517	56.8M
YOLOv12l	0.564	0.657	0.531	26.3M
YOLOv8l	0.559	0.673	0.547	43.6M

Only 5 of 14 classes had val instances — 9 classes had zero ground-truth boxes in this split

Tools-Detection-Rate on 60 Videos-Sessions

Tool	YOLOv8l	RT-DETR	YOLOv11x	YOLOv12l
Needle driver	57.0%	**72.0%**	64.7%	63.6%
Monopolar curved scissors	37.5%	**39.2%**	36.5%	**40.9%**
Bipolar forceps	21.9%	**27.8%**	22.9%	23.9%
Force bipolar	22.0%	**29.7%**	17.4%	15.4%
Cadiere forceps	18.8%	22.4%	21.0%	**53.0%**
Clip applier	3.8%	**10.9%**	2.6%	4.9%
Vessel sealer	0.1%	1.2%	0.03%	0.02%
Grasping retractor	0.1%	0.2%	0.2%	0.2%
Not in training data				
Stapler	0%	0%	0%	0%
Prograsp forceps	0%	0%	0%	0%
Cautery hook	0%	0%	0%	0%
Tip-up grasper	0%	0%	0%	0%
Suction irrigator	0%	0%	0%	0%
Bipolar dissector	0%	0%	0%	0%

Next Steps

- *In progress:*
  - 4-model evaluation across all 155 sessions (jobs running)
  - Manual annotation of extracted frames

- *Next:*
  - Retrain model with newly annotated classes
  - Find supplementary datasets for remaining missing classes
  - Begin classifier training