WEEK 3 UPDATE

SmartOR-Check & SurgVU26 Challenge

AI-Assisted OR Readiness · Surgical Tool Detection

DeAndre Bailey | Computer Science, UCF

Advisor: Dr. Chen Chen | Mentors: Yuanhao Zhou · Darinka Townsend Hinostroza

CRCV REU · Summer 2026

MY PROJECT

Two complementary AI challenges in the surgical domain

1 SmartOR-Check

AI-Assisted OR Readiness Verification

Goal: Answer readiness questions from OR images
Output: Human-readable readiness report
Approach: Detect objects → build scene record → answer checklist
Key question: Does detection-guided reasoning beat direct VLM prompting?

2 SurgVU26 Challenge

Surgical Tool Detection & Localization

Goal: Classify & localize 12 robotic da Vinci instruments
Output: Bounding boxes + class labels per frame
Challenge: Weakly supervised — only clip-level presence labels, very few ground-truth boxes
Strategy: Pseudo-labeling → fine-tuned YOLO/RT-DETR

MY PROJECT

The main focus for this week

SurgVU26 Challenge

Surgical Tool Detection & Localization

What does it mean to classify and localize?
Classifying and localizing means identifying what objects are present in an image and drawing bounding boxes around their locations.

What does weakly supervised mean?
Weakly supervised learning uses limited or less detailed labels (such as image-level labels instead of exact object locations) to train a model.

What is Pseudo-Labeling?
Pseudo-labeling is a technique where a model generates labels for unlabeled data and then uses those predicted labels as additional training data.

PROGRESS THIS WEEK

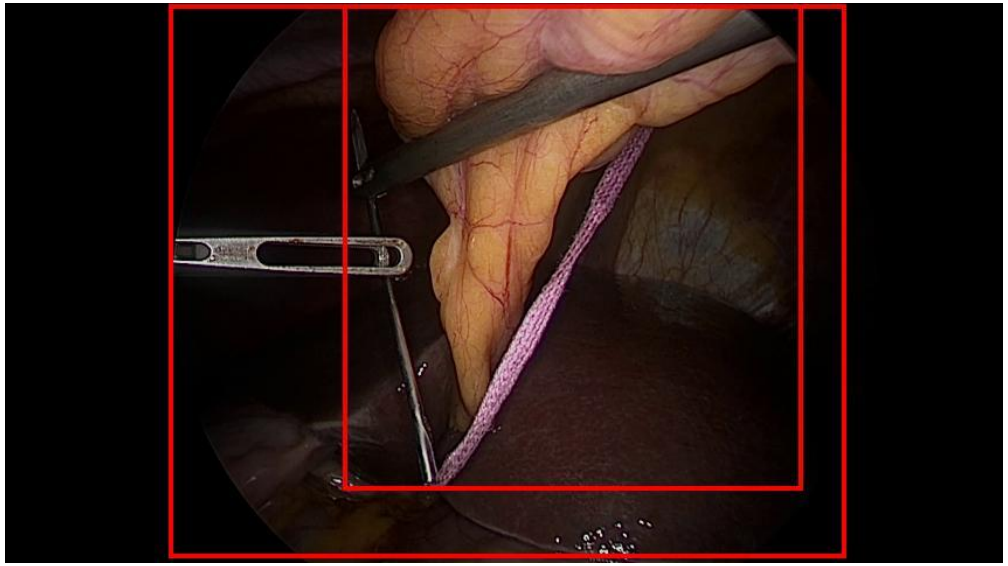
What I've worked on so far

📚 Literature & Challenge Review

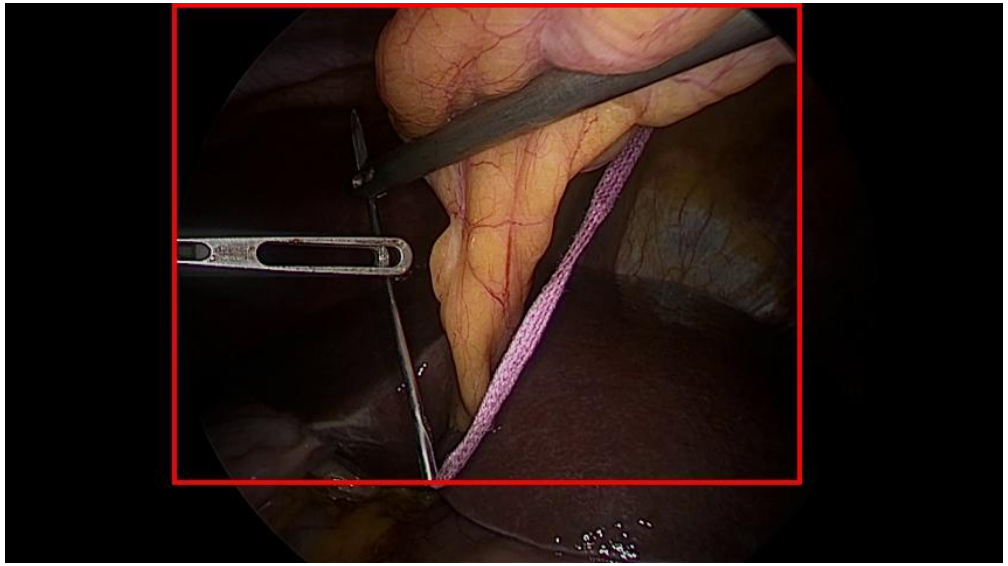
Read the SurgVU/SurcgToolLoc challenge papers (2022–2025). Analyzed what winning teams did: YOLO + pseudo-labeling + manual correction consistently outperform zero-shot VLMs.

🤖 Tested LocateAnything-3B (NVIDIA)

Set up & ran NVIDIA's new open-vocabulary grounding model on CholecTrack20 frames. Result: strong domain gap — the model localizes poorly on surgical footage without fine-tuning. This confirms pseudo-labeling alone won't work without a fine-tuned base.



LocateAnything Output Example



Grounding Dino Output Example

PROGRESS THIS WEEK

What I've worked on so far

Newton GPU Cluster Setup

Configured the UCF Newton HPC environment: built Conda env, pre-cached model weights (offline HuggingFace), and wrote Slurm batch scripts for testing models and YOLO Converter.

Potential Two-Stage Pseudo-Labeling

Designed an architecture: LocateAnything / YOLO → rough boxes → SAM 3 Tracker → precise masks → tight bounding boxes. Tested SAM 3's text vs. visual prompting; visual prompts are far more reliable for surgical tools.

Dataset Acquisition & Inspection

Downloaded CholecTrack20 frames (VID110). Confirmed bbox format (top-left normalized). Began writing YOLO converter. Evaluated OR-Arena dataset — no usable bounding boxes for our task. EndoVis17 Dataset has been added to Newton. Next step is to get SurgVU26 Dataset once approved for the challenge.

KEY FINDINGS

What the baseline tests revealed

01 Zero-shot VLMs fail on surgical imagery

LocateAnything-3B returned a single whole-frame box instead of finding individual instruments. The domain gap between everyday images and surgical endoscopy is severe — confirming that fine-tuning is required.

02 Pseudo-labeling needs a fine-tuned base

Raw VLMs can't pseudo-label effectively because their boxes are inaccurate. The winning SurgVU25 strategy: train a YOLO on the small validated box set first, then use it to generate pseudo-labels on the large weakly-labeled data.

03 Two-stage pipeline improves box tightness

LocateAnything → rough box → SAM 3 Tracker → pixel mask → tight box. SAM 3's visual prompting mode reliably refines rough boxes into clevis-accurate annotations, which are closer to what SurgVU actually scores.

04 2025 winners: YOLO + label quality > architecture

1st place (0.52 mAP) used Co-DETR + RTMDet. 2nd place (0.50) used YOLOv8 with manual label correction. The gap between 2nd and 3rd was data quality, not model size. A well-labeled YOLO beats a poorly-labeled transformer.

PLAN FOR THIS WEEK

Week 3 priorities — building toward a working baseline detector

1

Complete SurgVU → YOLO Converter

Finish the annotation converter (top-left → center normalized format) and set up the train/val split. This is the data pipeline foundation everything else needs.

DATA

2

Fine-tune First YOLOv8 Detector

Train YOLOv8 on CholecTrack20 as a proof-of-concept baseline. Goal: get first real mAP numbers and verify the full train → evaluate pipeline works before touching larger datasets.

TRAINING

3

Download EndoVis17 & SurgToolLoc Data

Request access to EndoVis17 (grand-challenge.org) and download SurgToolLoc training data from Google Cloud. These are the most critical datasets for da Vinci instrument detection.

DATASETS

4

Run LocateAnything / SAM 3 Comparison

Run the prompt evaluation script across 20 validation frames. Score each model against ground-truth boxes (IoU > 0.5). Determine whether the LocateAnything → SAM 3 two-stage pipeline gives meaningfully tighter boxes.

EVALUATION

Thank You For Listening!