Integrating Human-Object Interaction and Gaze Analysis towards Performance Assessment in CCAT Advanced Mixed-Reality Training

AFRL

Kent Etherton, PhD[1], Katelyn Kay, MS[1], Divya Mereddy, MS[2], Ashwin Tudur Sadashiva, PhD[2], Marcos Quinones-Grueiro, PhD[2], Gautam Biswas, PhD[2], Richard Strilka, MD[3], Maj William Tyler Davis, MD[4], Lane Frasier, MD[3], Benjamin Goldberg, PhD[5]

[1]Air Force Research Laboratory, Dayton, OH
[2]Vanderbilt University, Nashville, TN
[3]University of Cincinnati College of Medicine, Cincinnati, OH
[4]United States Air Force 59th Medical Wing, San Antonio, TX
[5]United States Army DEVCOM Soldier Center

Introduction

- Critical Care Air Transport (CCAT) teams require high-fidelity training for missions involving high-volume patient transports.
- Traditional evaluation methods are time-consuming and not scalable, limiting training throughput.
- This project introduces a human-AI assessment framework using computer vision and deep learning to provide objective, continuous evaluation of team performance during complex simulations.

Method

- We analyzed video data from 18 advanced CCAT simulation training sessions, which included ~103 million frames.
- Developed and fine-tuned Human-Object Interaction (HOI) model to detect and track trainee interactions with IVs, ventilators, and patients.
- To overcome visual occlusions, interaction predictions from different camera angles were merged and smoothed into continuous timelines.

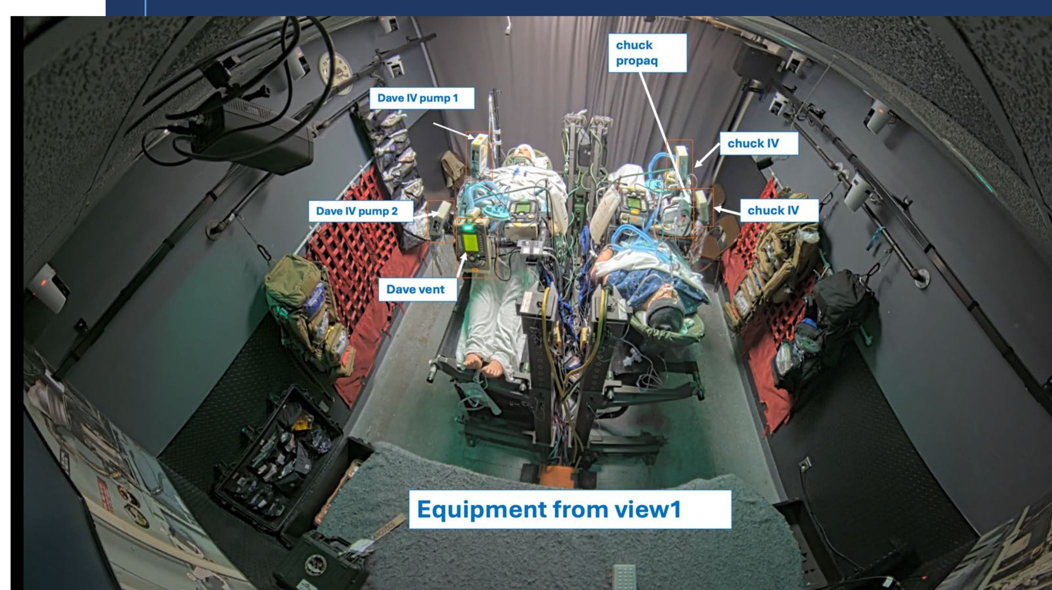

Figure 1: CCAT Advanced SICU Simulation Environment
Photo Credit: USAF School of Aerospace Medicine (USAFSAM), Air Force Research Laboratory, Dayton, OH

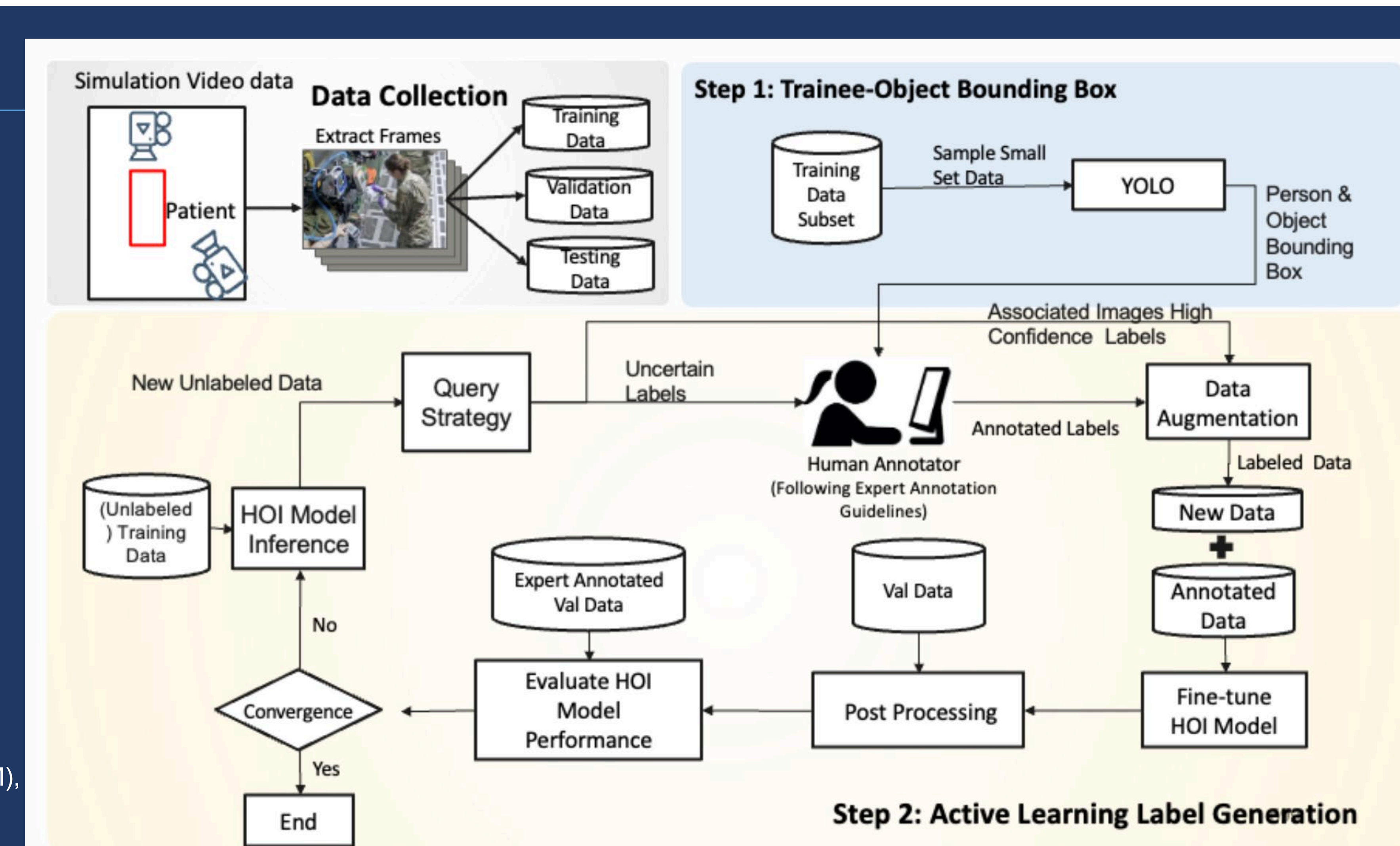

Figure 2: Active-learning-based iterative human-object interactions model finetuning

To efficiently train the Human-Object Interaction (HOI) model while minimizing manual effort, we implemented a **semi-automated active learning pipeline**:

Step 1 (Automated Pseudo-Labeling): We generated highly reliable initial training data by only accepting interaction labels where two independent AI systems (a pretrained HOI model and a hand-pose estimator) completely agreed.

Step 2 (Active Learning Iterations): The model analyzed new footage, automatically adding high-confidence predictions to the training set and flagging its most uncertain predictions for human review.

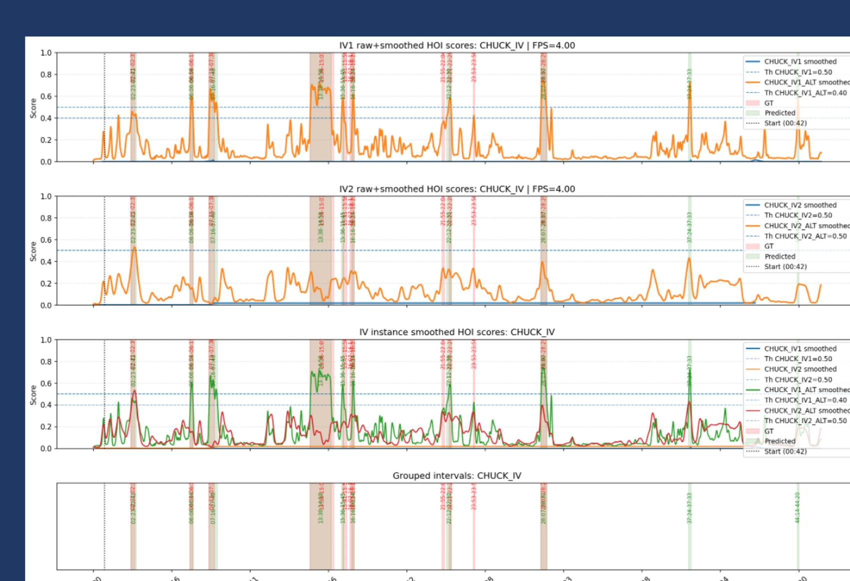

Figure 3: Two Object Family Equipment - IV Equipment (it includes two IV pumps)

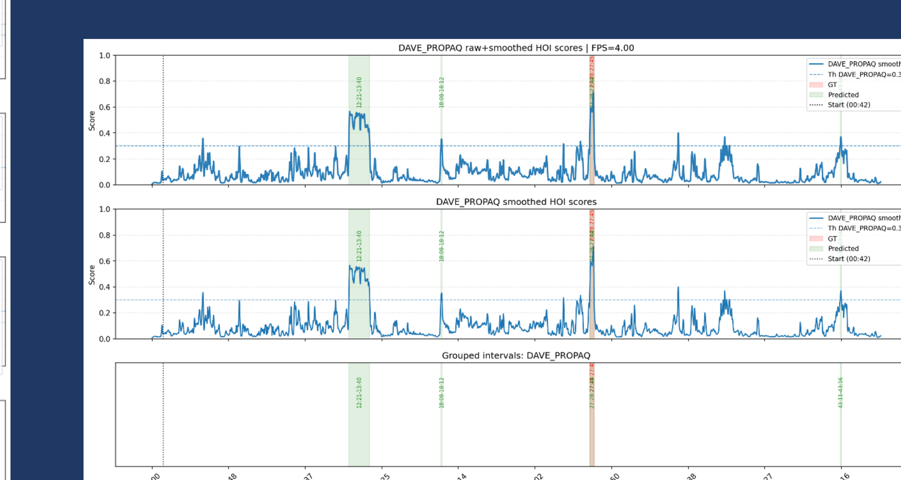

Figure 4: Propaq Interactions (Single Object Family Equipment)

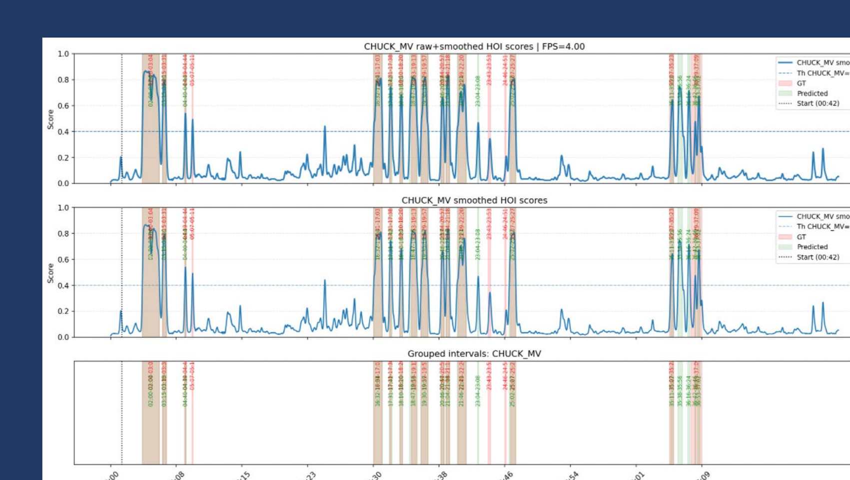

Figure 5: MV Ventilator (Single Object Family Equipment)

Results

The finetuned model achieved a **weighted F1-score of 0.97** on one complete labeled simulation, indicating reliable detection of clinically meaningful trainee-equipment interactions under realistic operational conditions. Detailed per-interaction weighted F1 scores are:

(Sim) Patient	Equipment interaction	Weighted F1
1	IV pump	0.954
1	Propaq monitor	0.982
1	Mechanical ventilator	0.994
2	IV pump	0.940
2	Propaq monitor	0.970
2	Mechanical ventilator	0.987

Conclusion and Future Work

We successfully trained computer vision models to recognize trainee-equipment interactions with high accuracy, demonstrating its utility when supplementing performance assessment.

Future work will incorporate gaze analysis using a trainee's estimated field of vision to infer situational awareness, measuring visual attention on critical equipment and alarms to better understand the cognitive processes behind a trainee's actions.

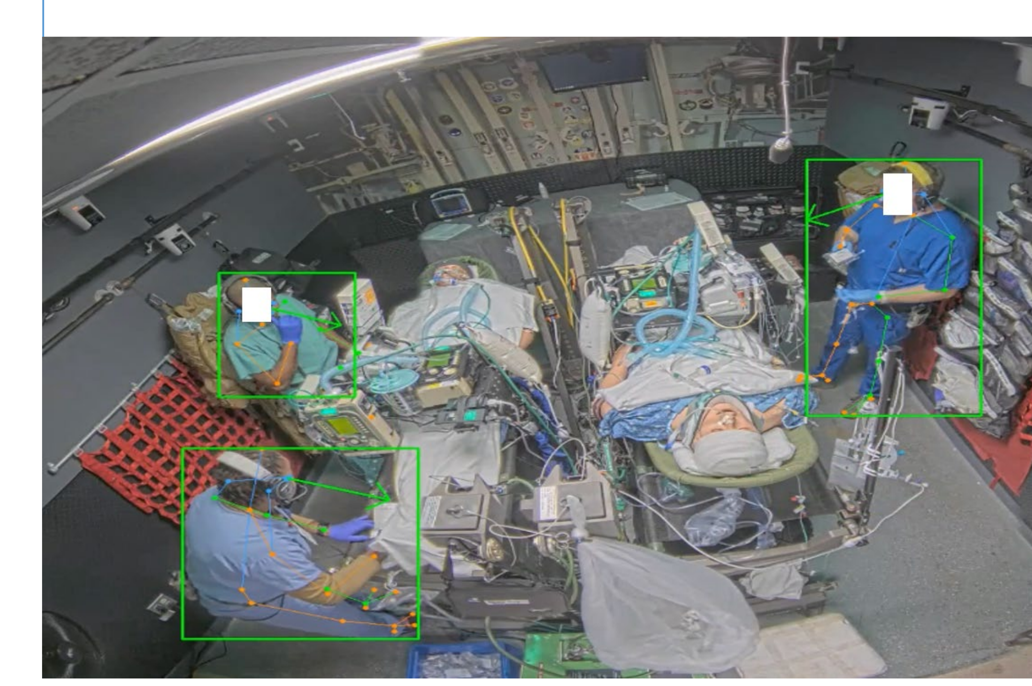

Figure 6: Preliminary gaze-vector and sectoral inferences
Photo Credit: USAF School of Aerospace Medicine (USAFSAM), Air Force Research Laboratory, Dayton, OH

THE AIR FORCE RESEARCH LABORATORY