

## Motivation

**Problem:** Although state-of-the-art (SOTA) HOI detectors perform well on benchmark datasets, their performance often degrades when deployed in real-world training environments due to domain-specific objects, occlusions, and complex visual conditions.

**Gap:** Prior HOI diagnosis studies have been developed primarily in the context of benchmark datasets. Systematic error-factor attribution frameworks remain relatively underexplored in the HOI literature

## CCATT Mixed-Reality Medical Training

**Environment:** Critical Care Air Transport Team Initials — 3-member team (physician, nurse, respiratory therapist) managing critically injured patients during aeromedical evacuation.

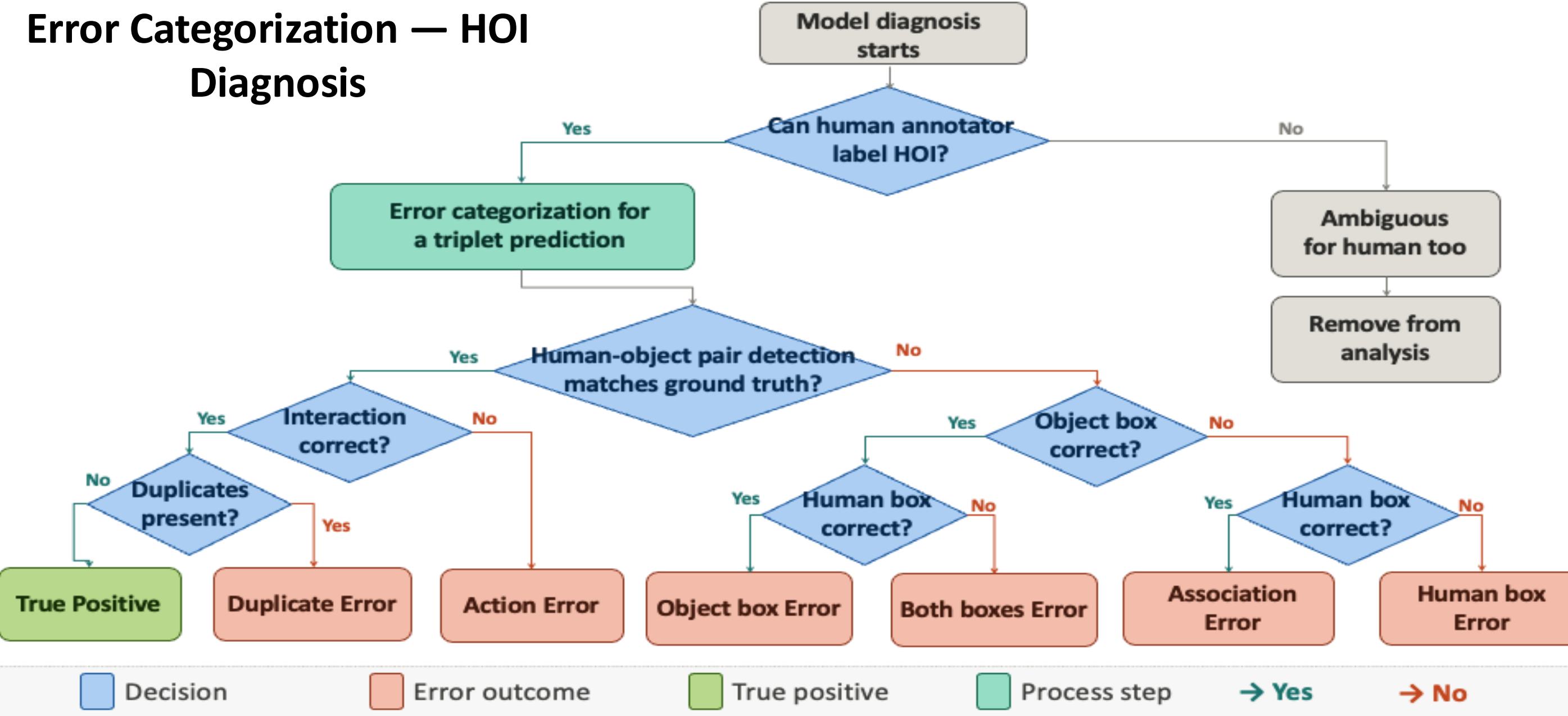
**Why HOI Matters:** HOI detections serve as measurable proxies for trainee competency — enabling an automated trainee evaluation system.

**Data:** 20 training simulations, multiple cameras, ~3.5 million frames (~45 hours of data sampled at 4 FPS).

## Contributions

- **Diagnostic framework for HOI models:** Triplet-level error taxonomy integrated with error-factor attributes to analyze HOI failure modes under real-world deployment conditions.
- **Diagnosis-guided refinement strategies:** Mitigation strategies selected directly from diagnosed error causes — automated augmentation + human-in-the-loop correction — demonstrated in the CCATT setting.
- **Error analysis on CCATT mixed-reality data:** Pretrained CDN, PVIC, and HOICLIP evaluated on en-route critical care training video data — diagnostic insights directly informed CDN model adaptation.

## Error Analysis Framework



## HOI Error Category with Co-occurring Error Factors

HOI Error Bucket	Visual Degradation Factors					Domain Mismatch			Data Factors		Pretrained Model
	Low-light /IR	Blur	Semi-Occl.	Camera angle	Crowded scenes	Objects	Actions	Humans	Imbalance	Verb. ambig.	No-int. sparsity
Detection – Object Box Error	✓	✓	✓	✓	✓	✓			✓		✓
Detection – Human Box Error	✓	✓	✓	✓	✓			✓			✓
Detection – Both Boxes Error	✓	✓	✓	✓	✓	✓					✓
Association Error		✓	✓	✓	✓						
Action Error	✓	✓	✓	✓			✓		✓	✓	✓

Legend: ✓ Automated refinement, ✓ Human annotator-in-the-loop, ✓ Not addressable, No co-occurrence

## HOI Model Diagnosis: Refinement Strategies

Error Factor	Refinement	Refinement Type
Blur	Blur augmentation (motion + Gaussian)	Automated
Low-light / IR	Grayscale and illumination variation augmentation	Automated
Occlusion	Partial occlusion masking during training	Automated
Crowded scenes	Human-in-the-loop annotation correction	Human
Camera-angle	Human-in-the-loop annotation correction	Human
Domain mismatch – Objects / Person	Add CCATT objects' bounding boxes using YOLO	Automated
Domain mismatch – Actions	Add CCATT action labels & human annotations	Human
Class imbalance	Focal loss & minority class augmentation	Automated
No-interaction	Add explicit no-interaction annotation	Automated

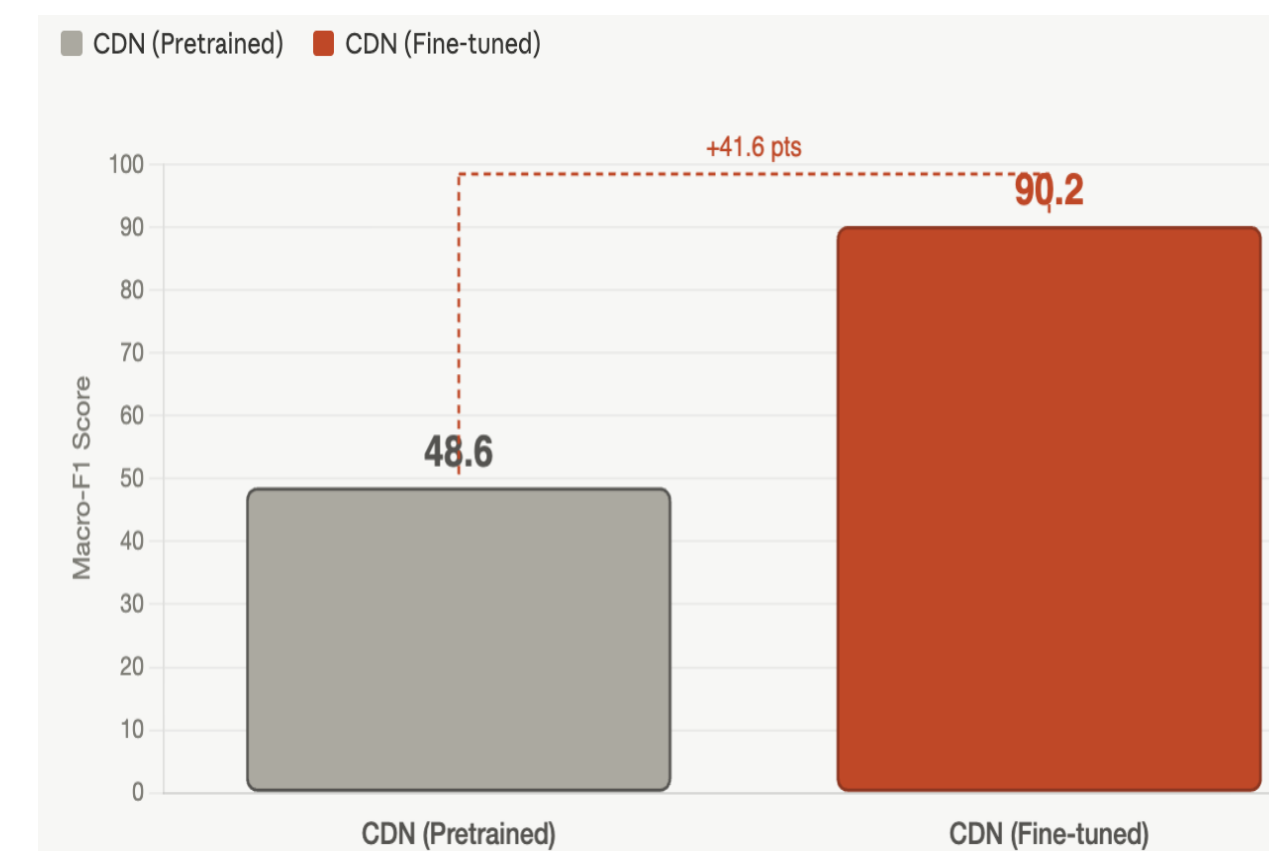
## Distribution of HOI error buckets and associated co-occurring error factors across pretrained detectors

HOI Errors	Co-occurring Error Factors	CDN	PVIC	HOICLIP
Detection Object Box Error	Domain Mismatch	35.50	38.45	2.10
	Classification Error – Domain Mismatch	100.00	100.00	100.00
	Semi-occlusions	3.99	3.36	0.84
	Low-light / IR	38.24	38.45	2.10
Detection Person Box Error	Domain Mismatch	1.05	1.05	5.25
	Low-light / IR	1.05	1.05	2.52
	Semi-occlusions	1.05	0.00	2.94
Association Error	Crowded Scenes	0.84	2.10	6.09
	Low-light / IR	0.63	2.10	6.30
Action Error	Domain Mismatch	23.74	21.01	8.61
	Classification Error – Domain Mismatch	12.61	17.86	22.48
	Blurred Regions	0.42	0.21	0.42
	Low-light / IR	13.03	17.86	8.82

Access the full paper here



## Diagnosis Results – Performance Improvement



## Human–Object Interaction (HOI) Error Categories: Qualitative Examples of CCATT-Specific Errors (AFRL, Dayton)

