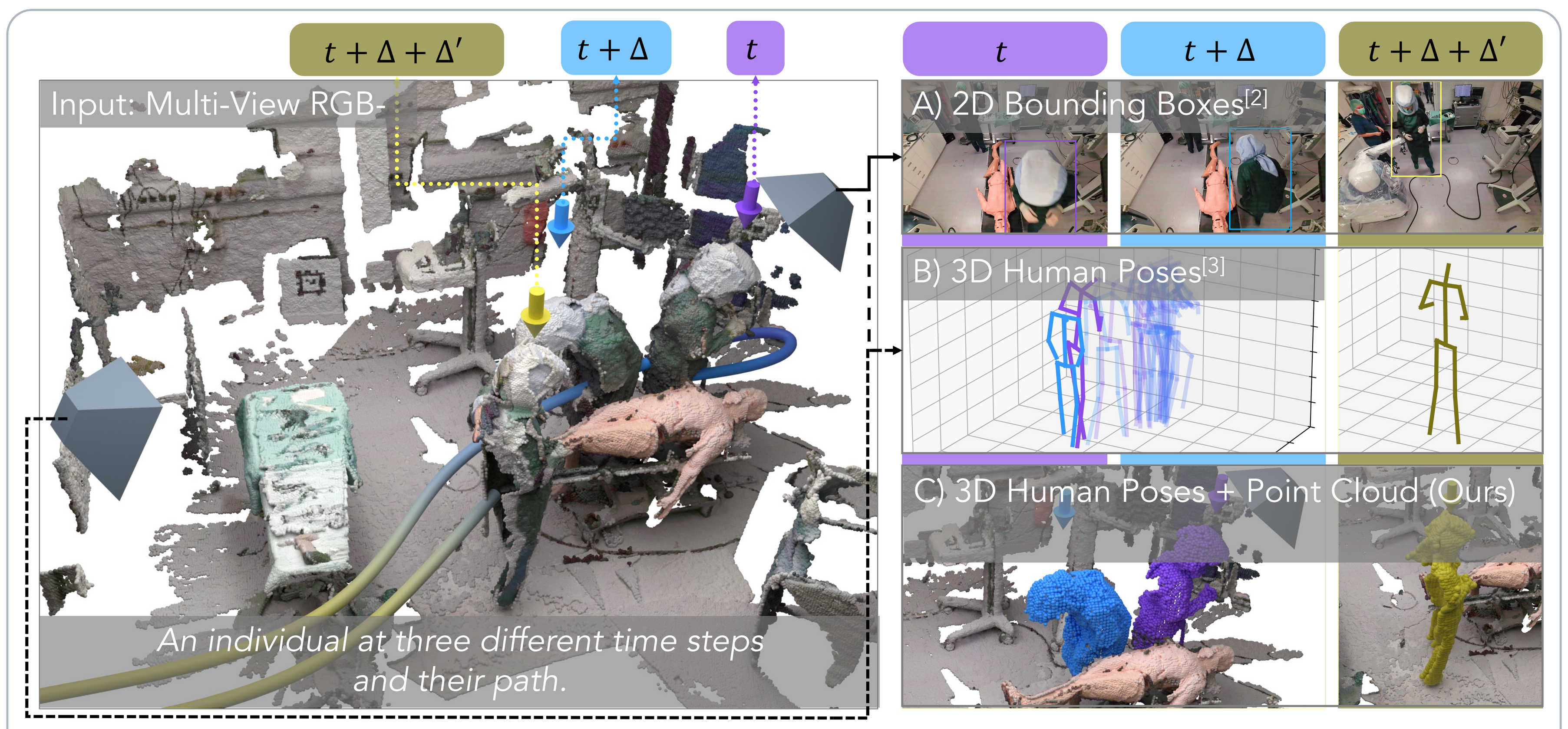
TrackOR: Towards Personalized Intelligent Operating Rooms Through Robust Tracking

Tony D. Wang[1], Christian Heiliger[2], Nassir Navab[1,3], Lennart Bastian[1,3]

[1]Computer Aided Medical Procedures, Technical University of Munich, Germany
[2]Minimally Invasive Surgery, University Hospital of Munich (LMU), Germany
[3]Munich Center for Machine Learning, Germany

Introduction

- The next generation of intelligent systems in the OR requires a fundamental shift from **generic, role-based analysis to a personalized, staff-centric** understanding of the surgical team
- This shift is impeded by the **OR's visually homogeneous attire** and **frequent staff absences** that confound traditional, appearance-based tracking and re-identification (ReID) methods.
- We propose ***TrackOR***, a framework that uses robust, view-invariant 3D geometric signatures[1] for ReID , enabling the **persistent, long-term tracking** required for granular workflow analysis

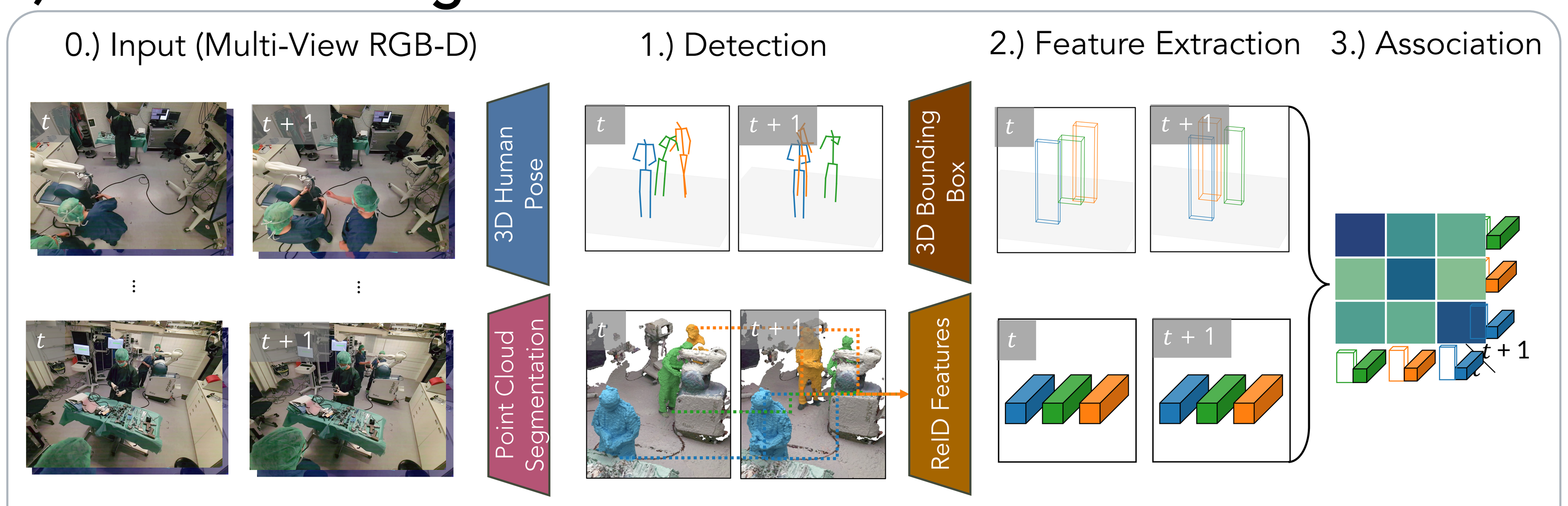


- **A) 2D Trackers[2]:** Fail due to single-camera views and ambiguous, appearance-based ReID.
- **B) Standard 3D Trackers[3]:** Handle multi-camera setups but lack robust ReID features, so they cannot track staff through prolonged absences.
- **C) Our Method:** Creates strong **geometric ReID features**, enabling persistent identity tracking even when staff leave and re-enter the OR.

Methodology

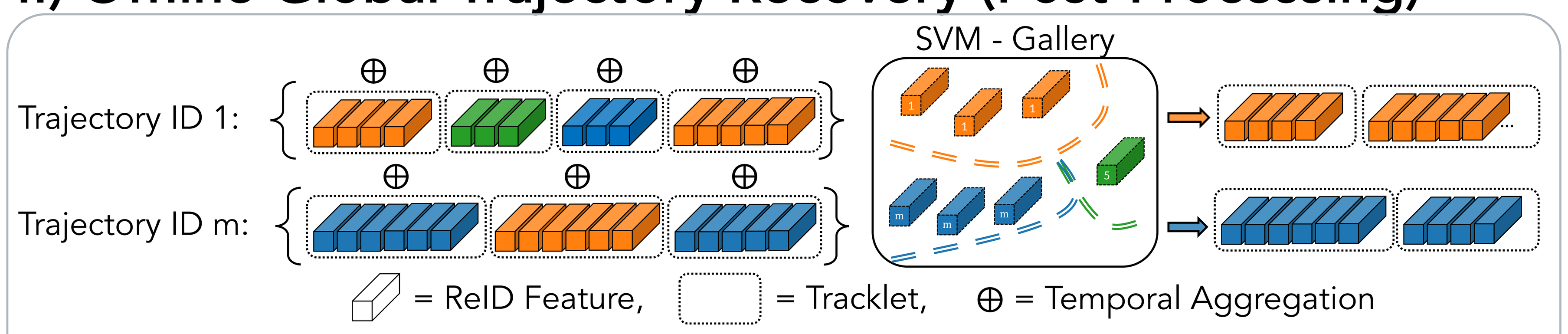
Overall Approach: *TrackOR* is a "tracking-by-detection" framework. Its core is a **real-time online tracker** that achieves SOTA performance. For downstream applications requiring fully reconstructed paths, the framework also includes an optional **offline recovery process.**

I) Online Tracking



1.) 3D Detection: We take a "3D-first" approach, detecting human poses directly in 3D from multi-view camera inputs.
2.) Feature Extraction: A robust, view-invariant 3D geometric signature is extracted from each person's point cloud to serve as the ReID feature.
3.) Association: Detections are matched to existing trajectories using a cost matrix that combines a **spatial cost** (3D GIoU) with a **shape cost** based on the cosine dissimilarity of our geometric ReID features.

II) Offline Global Trajectory Recovery (Post-Processing)



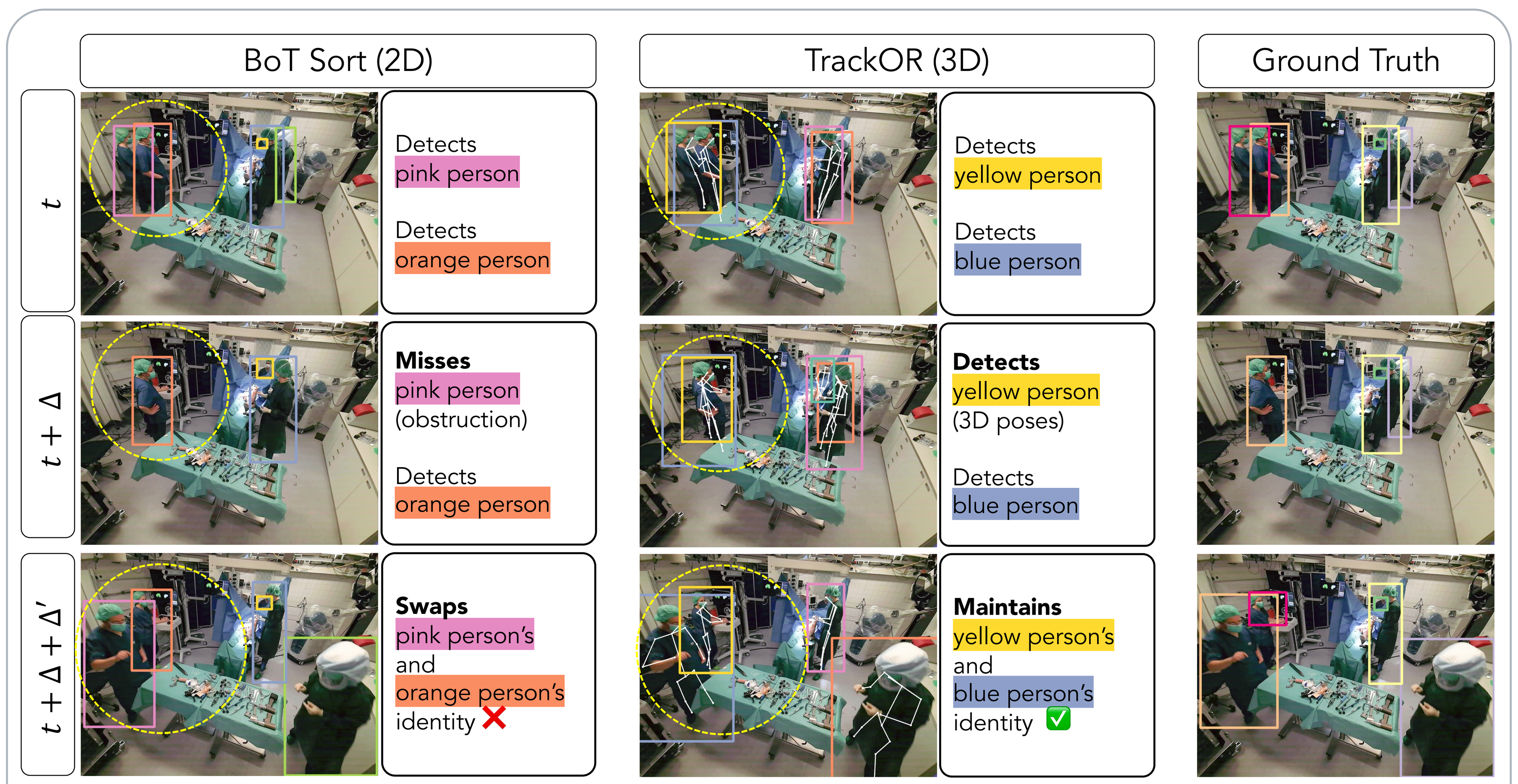
This post-processing step can be applied for specific analyses, such as our *temporal pathway imprints.*
- An SVM-Gallery classifies each tracklet by identity.
- All tracklets assigned the same identity are then merged to form a complete, persistent trajectory for each individual.

Experiments & Results

Experimental Setup: We evaluate our method on the challenging **MM-OR dataset**[4], which features **frequent occlusions** and **visually homogeneous attire**. Performance is measured using the standard **Higher Order Tracking Accuracy** (HOTA) metric and **Multiple Object Tracking Accuracy** (MOTA).

Model		$HOTA_{(\alpha=.05-.5)}$ [17]			Identity[23]		CLEAR [7]			Count		Speed
Tracker	ReID	HOTA↑	AssA↑	DetA↑	IDF1↑	IDSW↓	MOTA↑	FP↓	FN↓	% #Dets	% #IDs	FPS↑
2D Bounding Box Tracker, using [28] as detections.												
OC-Sort [8]	✗	49.660	27.054	91.158	40.089	566	79.575	849	**566**	102.92	362.26	850
ByteTrack [30]	✗	58.430	37.451	91.163	52.705	312	75.946	950	1071	98.75	141.50	997
Strong Sort [11]	RGB	57.965	36.783	91.347	43.378	377	72.729	679	744	**99.33**	164.15	20
Boost Track [26]	RGB	54.848	33.013	91.130	42.989	511	77.420	622	1057	95.52	1,037.74	60
Deep OC-Sort [19]	RGB	78.359	66.348	**92.545**	73.007	200	**85.256**	**437**	793	97.36	239.62	34
BoT Sort [1]	RGB	80.825	71.309	91.612	74.686	266	78.936	837	940	98.94	139.62	32
3D Human Pose Tracker, using [24] as detections.												
KSP Tracker† [5]	✗	54.037	36.086	80.918	46.369	462	51.768	1566	2650	88.82	158.49	115
Nearest-Neighbor [21]	✗	73.366	65.813	81.787	66.496	86	55.686	1564	2648	88.82	113.21	2365
Kalman Filter [16]	✗	71.047	62.441	80.840	63.317	**80**	54.686	1543	2772	88.82	**101.88**	1121
TrackOR (Ours)	Depth	**82.216**	**82.300**	83.685	**76.362**	125	55.284	1564	2648	88.82	130.19	17

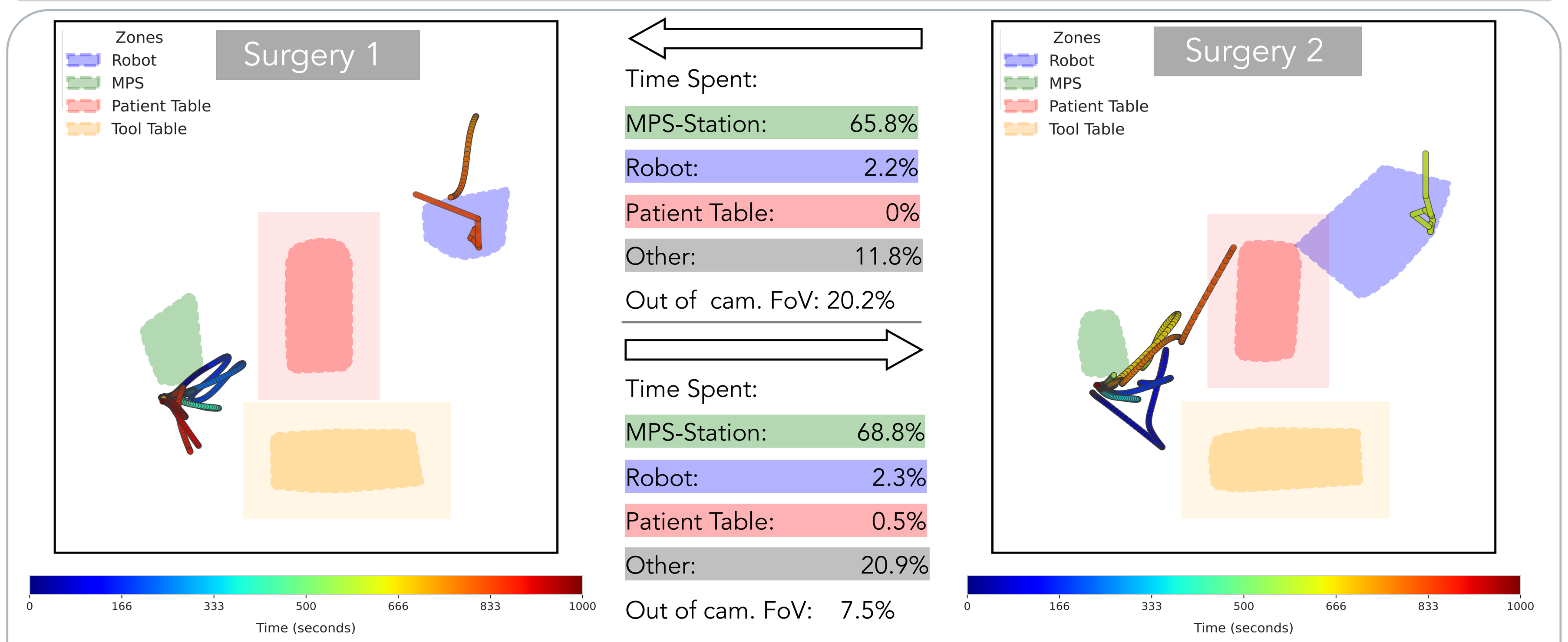
- TrackOR achieves a state-of-the-art **HOTA score of 82.2%**, outperforming all 2D and 3D baselines.
- Our performance is driven by a superior **Association Accuracy (AssA) of 82.3%**, highlighting TrackOR's ability to maintain correct identities.



- The figure above shows a common failure case where an appearance-based 2D tracker incorrectly **swaps identities** after a person is temporarily obstructed.
- In the same scenario, our 3D method successfully handles the obstruction and **maintains the correct identities.**

Temporal Pathway Imprints

As a **downstream application** of TrackOR, we introduce *temporal pathway imprints*. These leverage the complete, recovered trajectories to visualize an **individual's movement over time** from a bird's-eye view.



- The imprints reveal workflow differences for the same technician across two surgeries, showing different patterns of interaction with the robot.
- The Surgery 2 pathway shows the **non-sterile technician** coming into proximity with the **sterile patient table** – a potential safety concern.

References & Acknowledgments

[1] Wang et al., in Medical Image Analysis (2025)
[2] Aharon et al., in arXiv (2022)
[3] Liu et al., in Artificial Intelligence Surgery (2024)
[4] Özsoy et al., in Conference on Computer Vision and Pattern Recognition (2025)

Acknowledgements: This work was partly supported by the state of Bavaria through Bayerische Forschungsstiftung (BFS) under Grant AZ-1592-23-ForNeRo and the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through the Central Innovation Programme for SMEs (ZIM) under Grant KK 5389102BA3.