

CholecTrack20

A Multi-Perspective Tracking Dataset for Surgical Tools



Chinedu Nwoye^{1,3}



Kareem Elgohary¹



Anvita Srinivas¹



Fauzan Zaid¹



Joel Lavanchy²

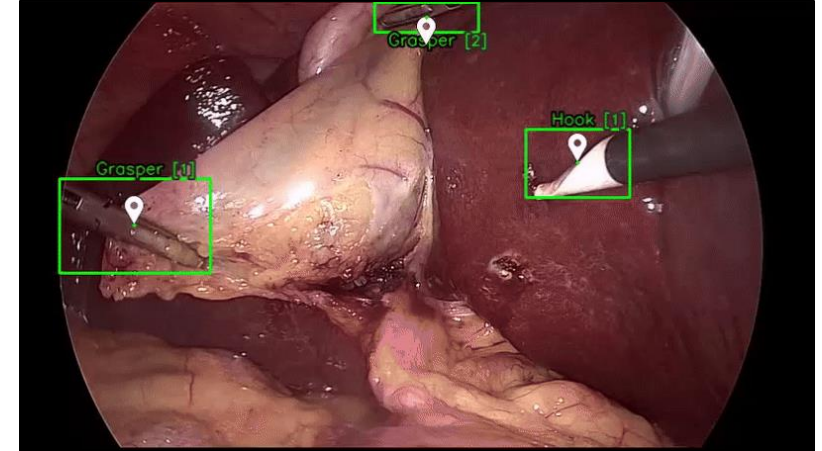
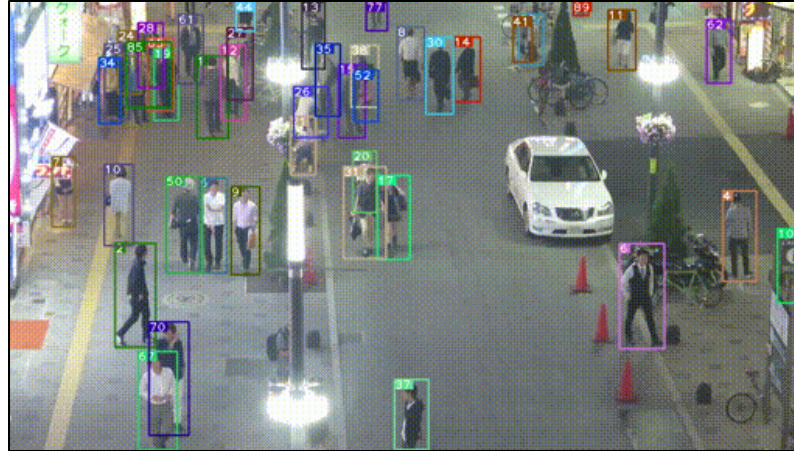
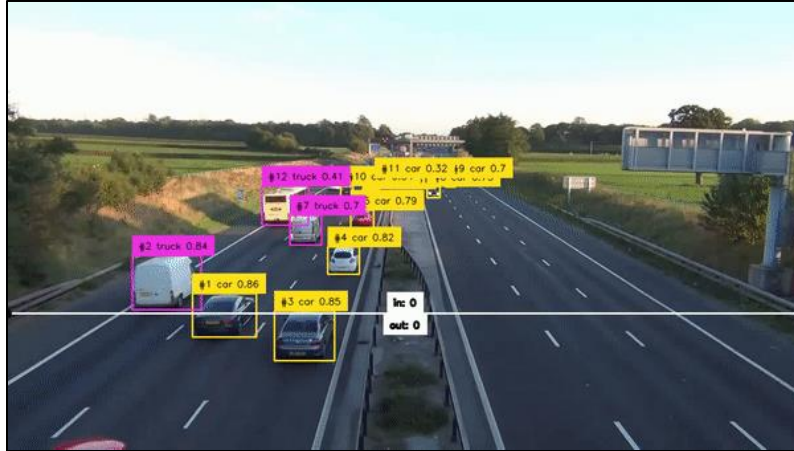


Nicolas Padoy^{1,3}

¹ University of Strasbourg, CNRS, INSERM, ICube, UMR7357, Strasbourg, France

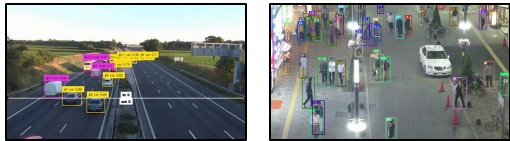
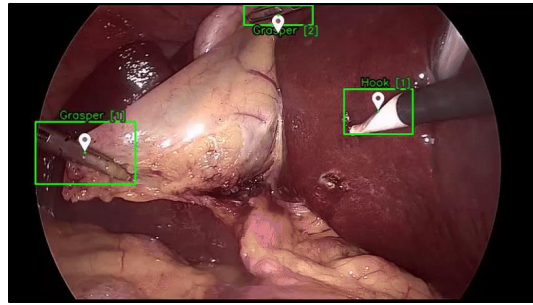
² University of Basel, University Digestive Health Care Center, Clarunis, Switzerland

³ IHU Strasbourg, France



- Object tracking estimates object location across time
- It enables perception, understanding, action, and analysis.

Surgical tool tracking is central to modeling surgical workflows and automation.



Why track surgical tool?

decision support

surgical skill
assessment

navigation

surgical safety

risk zone estimation

laparoscope positioning

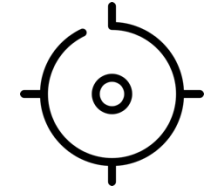
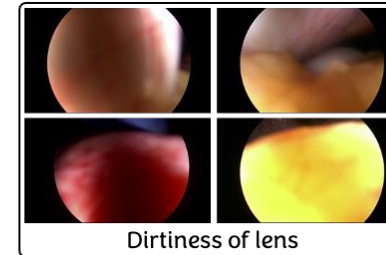
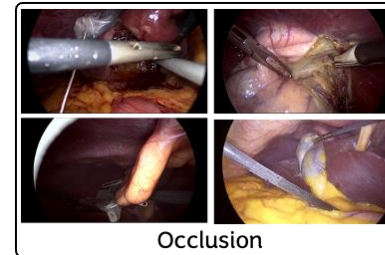
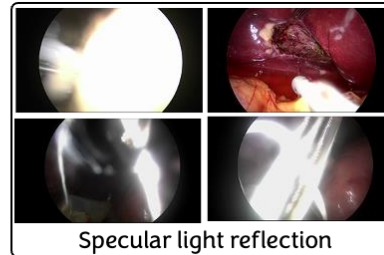
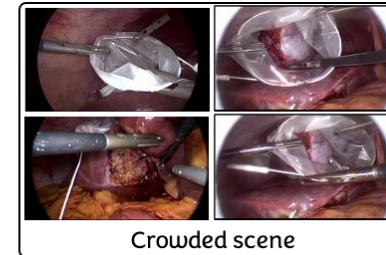
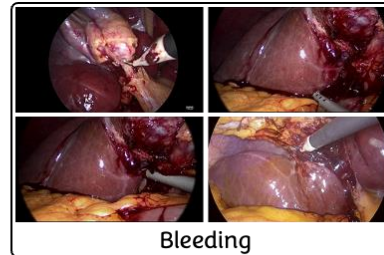
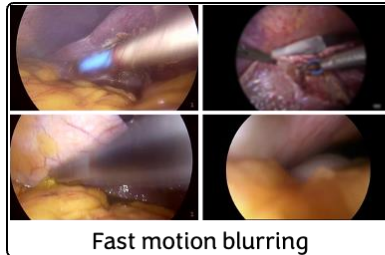
robotic visual servoing

insights

virtual and augmented reality

- Object tracking estimates object location across time
- It enables perception, understanding, action, and analysis.

Surgical tool tracking is central to modeling surgical workflows and automation.



Surgical scene is **dynamic and constrained**

Tools follow **non-rigid, disappearing trajectories**

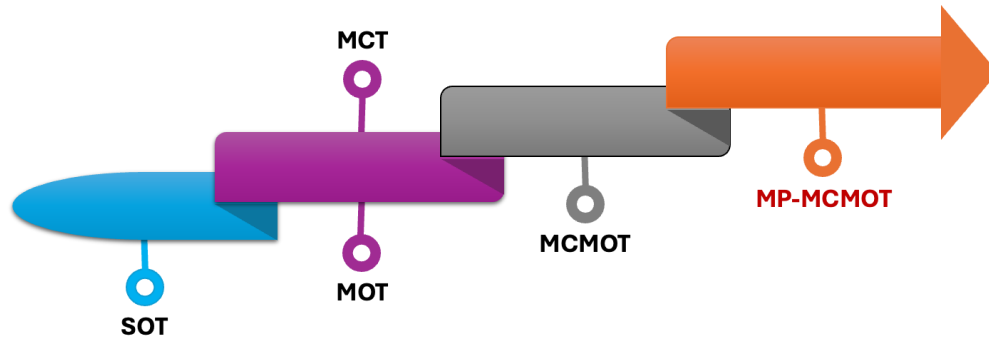
Multiple tools are **visually similar and used simultaneously**

Occlusion is frequent – tissue blocks and deformable, 2D monocular view

Tools go **out-of-camera view, out-of-body** and later **reappear**

Tool's appearance can change by **stains from blood, fluid, or light reflection**

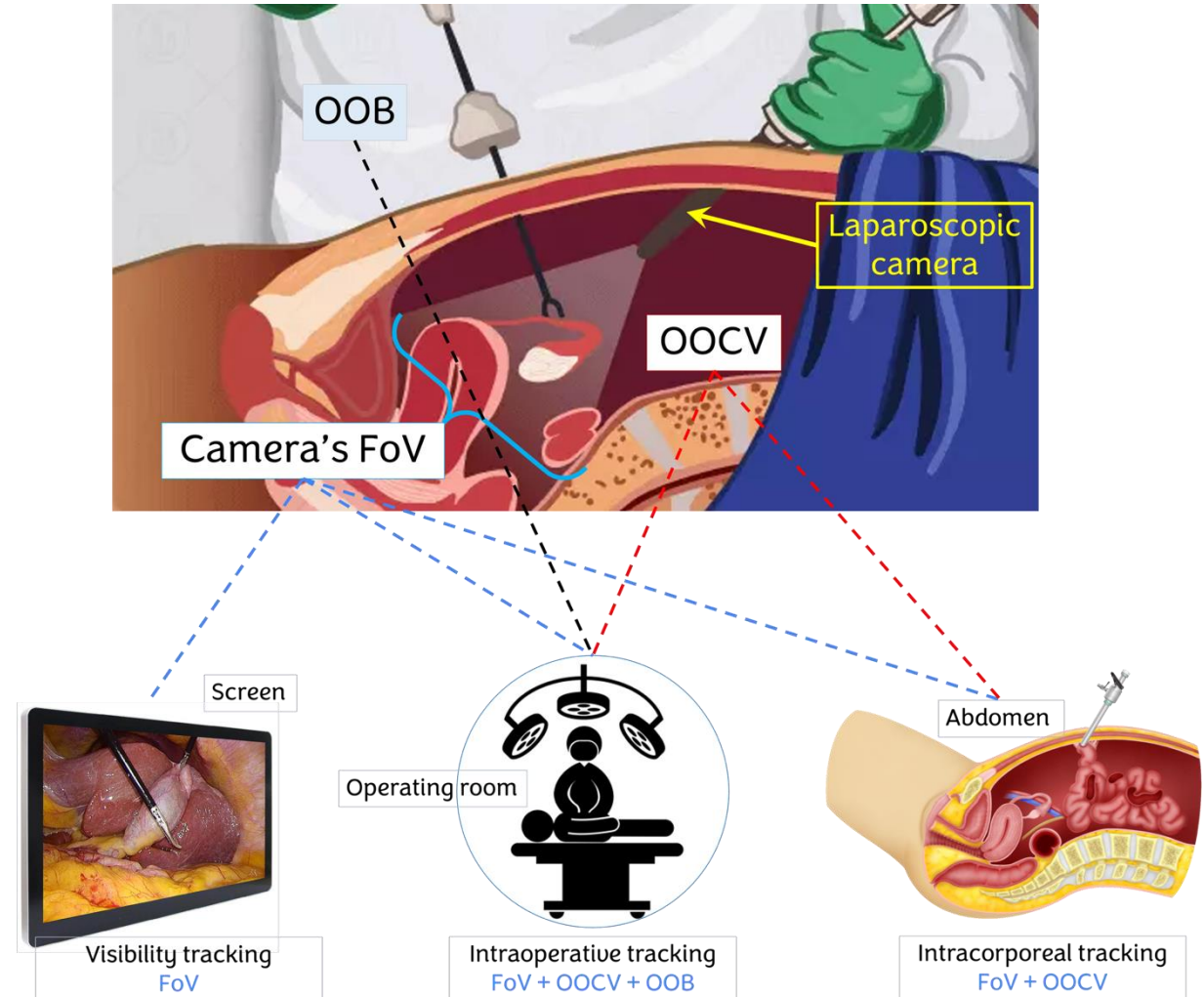
- Existing tracking **fail** to capture the **complexity** of surgical scene.



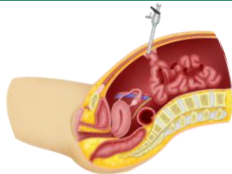
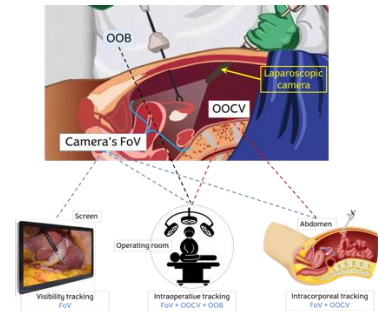
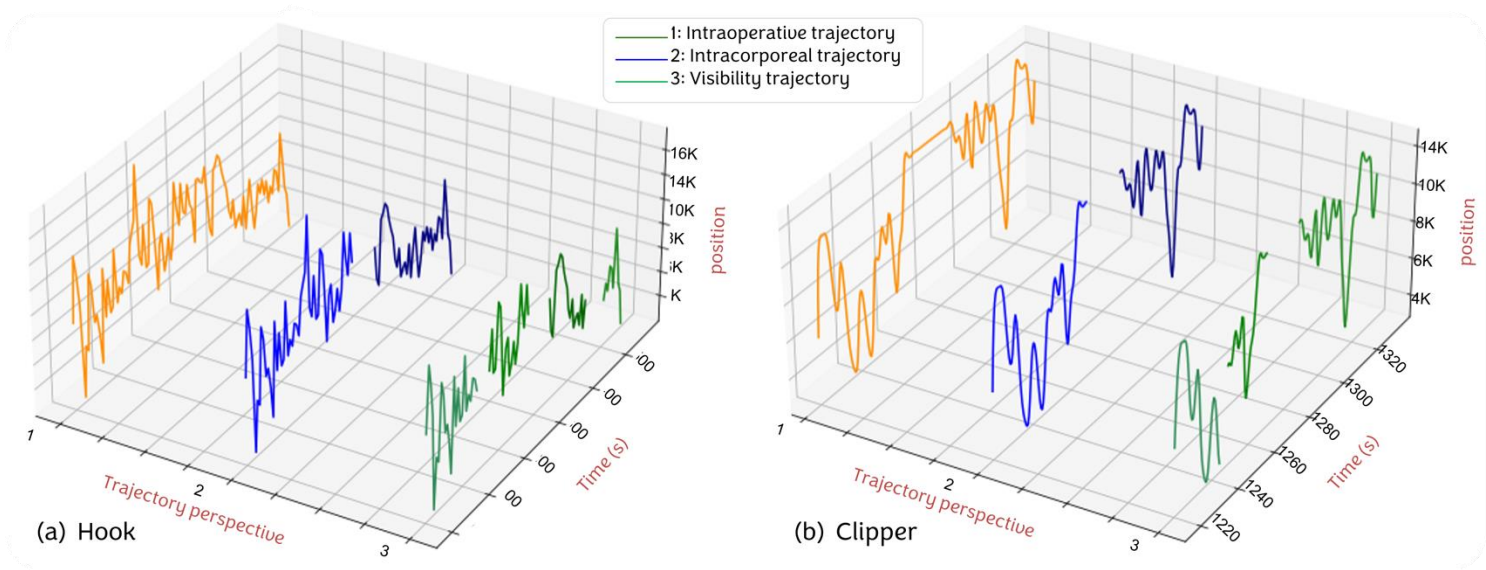
Tools move across views:


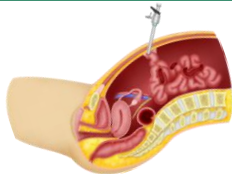

- operating room / out-of-body
- abdomen / within-the-body
- camera field of view / on-screen

➤ **track identity becomes ambiguous.**

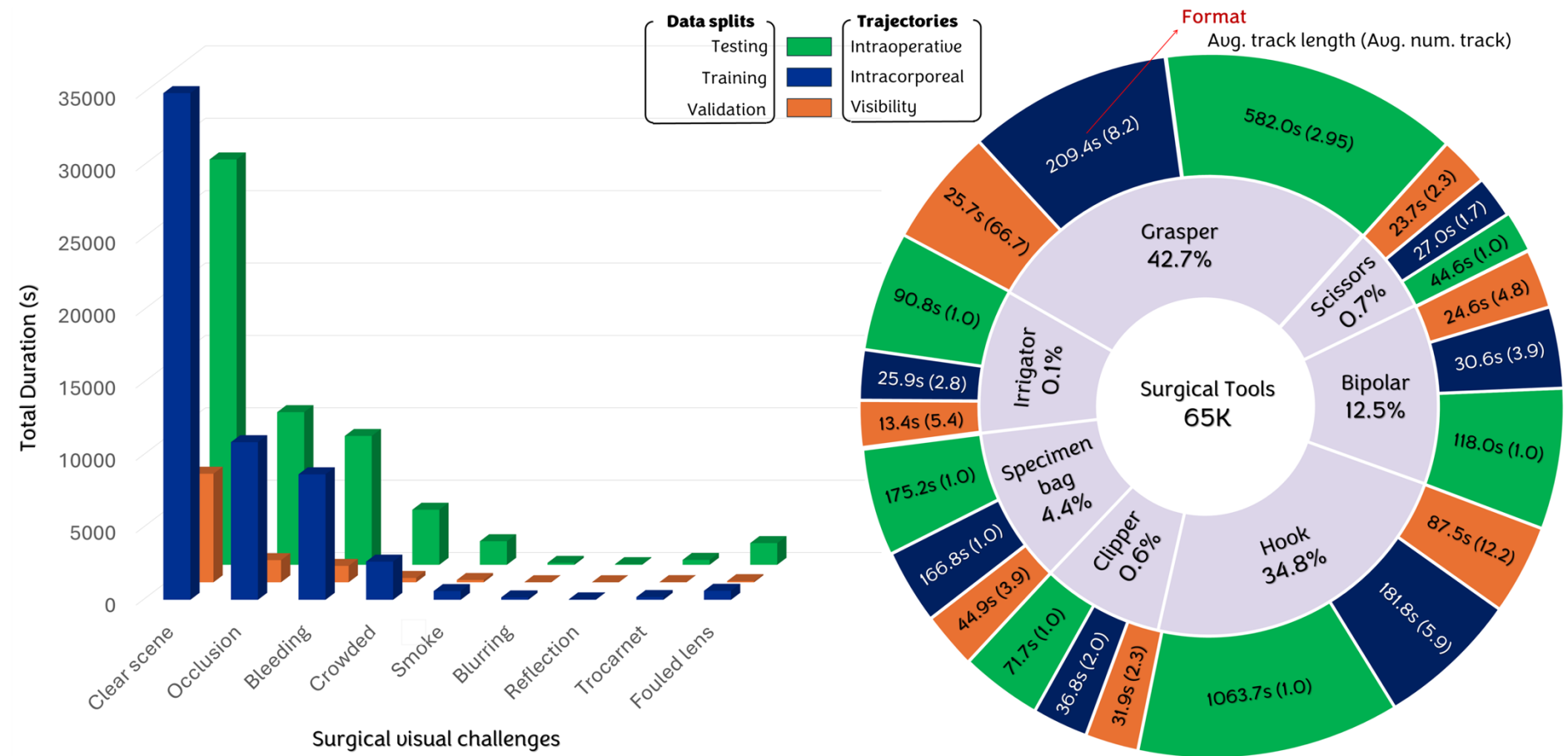


- Multi-perspective tracking paradigm:
 - redefines tracking as perspective-aware trajectories.
 - 3 tracking perspectives:



<div>intraoperative</div> <div></div>	tool life-long usage timeline	
	<div>intracorporeal</div> <div></div>	tool inside the body timeline
	<div>visibility</div> <div></div>	tool in camera focus timeline

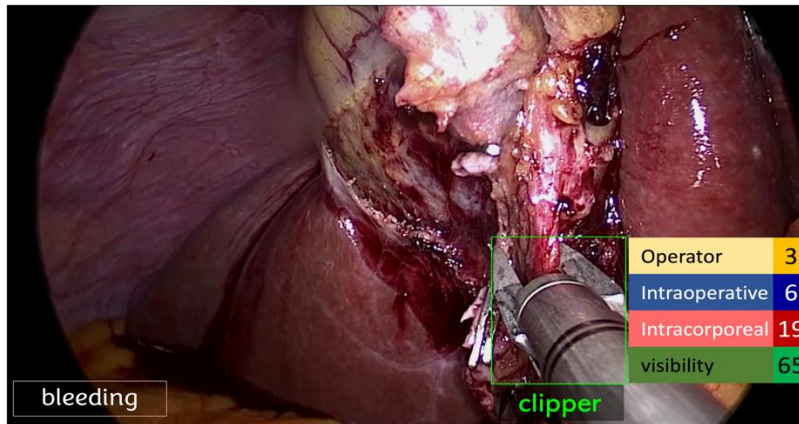
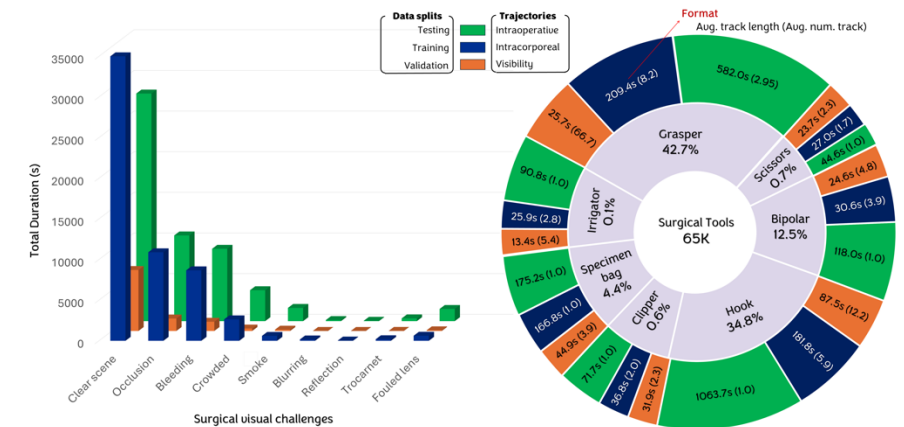
- First multi-perspective surgical tracking dataset
- 20 videos → 35K frames → 65K tool instances



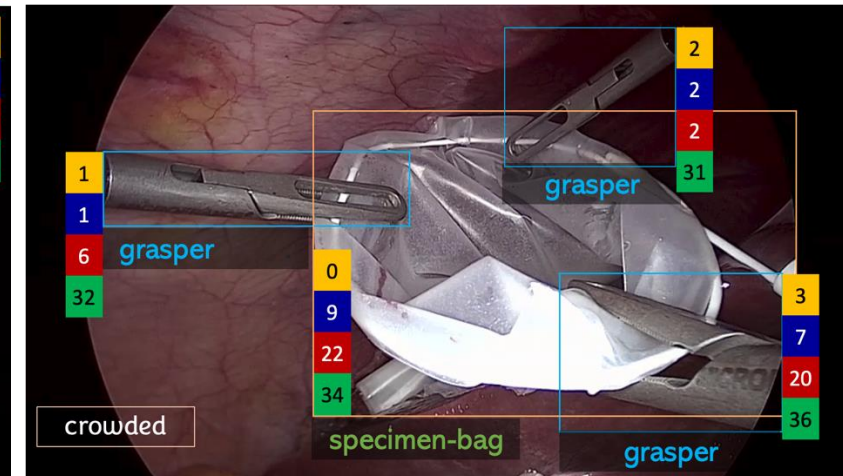
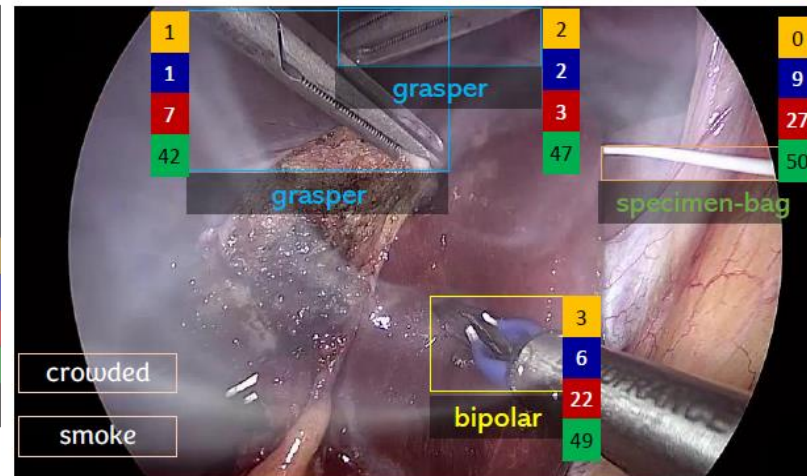
- First multi-perspective surgical tracking dataset
- 20 videos → 35K frames → 65K tool instances

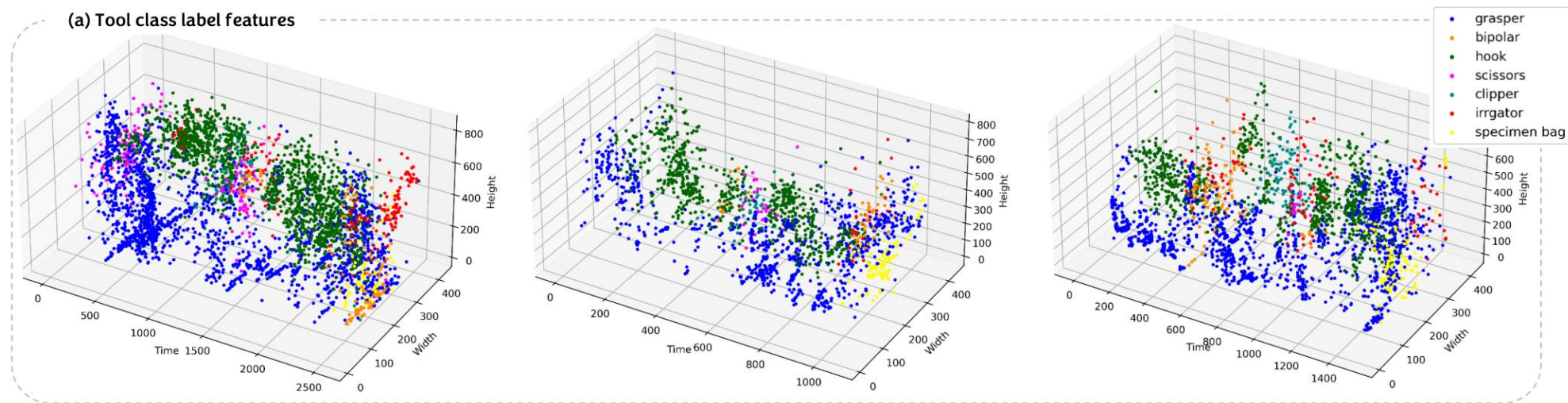
Annotations include:

- Spatial locations (bounding boxes)
- Tool categories
- Tool identity across 3 perspectives
- Surgeon operating hand identities
- Surgical phases
- Visual challenges (e.g. smoke, bleeding, ...)

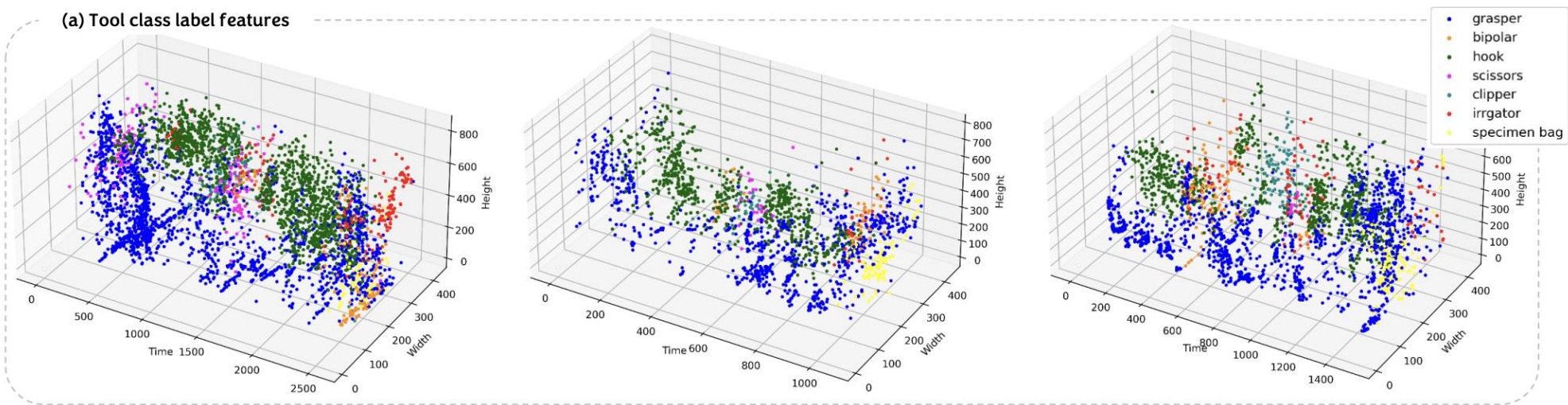


Phases: ■ ■ ■ ■ Clipping & cutting

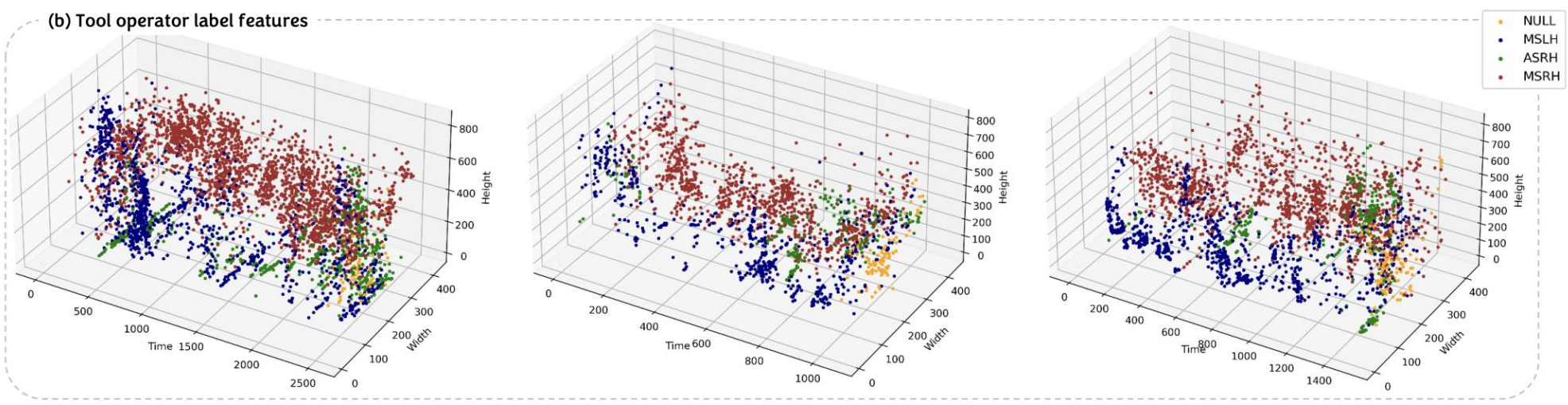


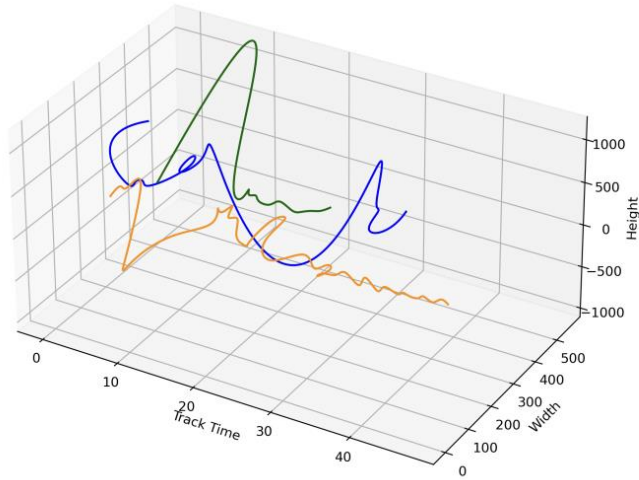


(a) Tool class label features

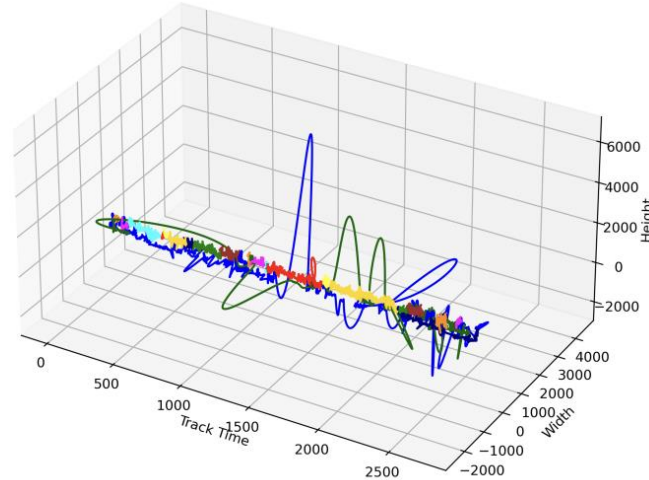


(b) Tool operator label features

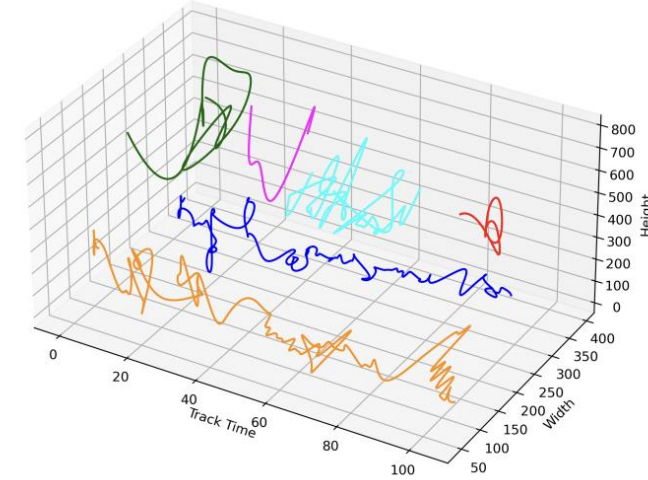




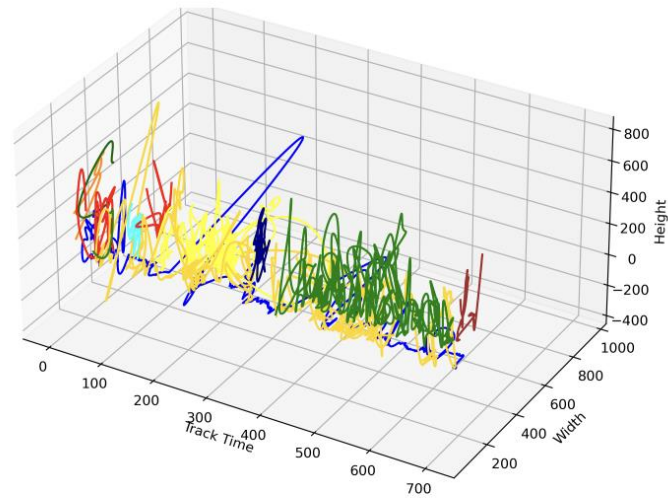
(a) Preparation



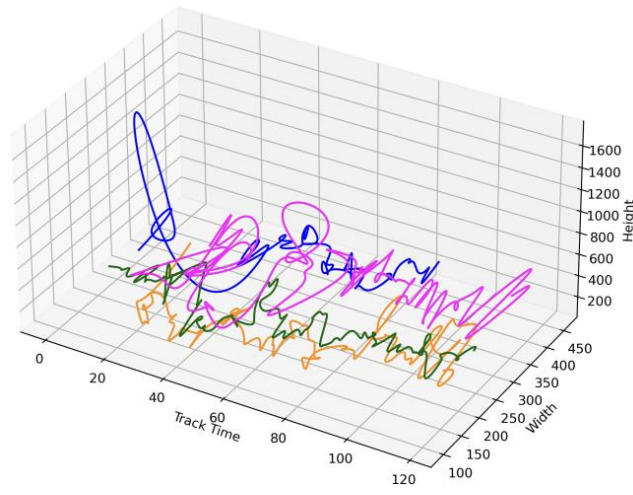
(b) Calot triangle dissection



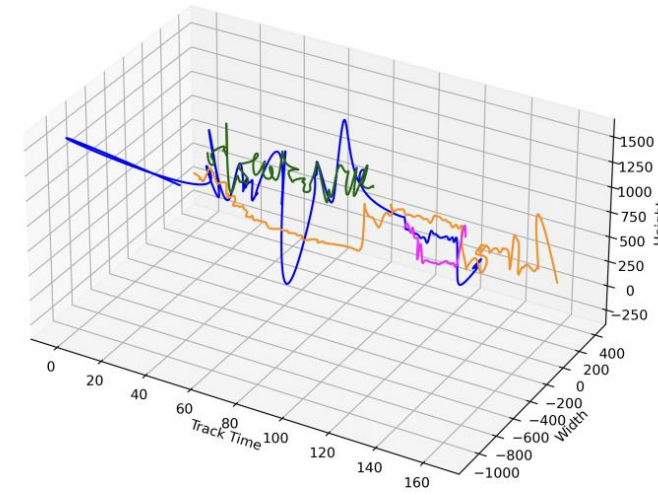
(c) Clipping and cutting



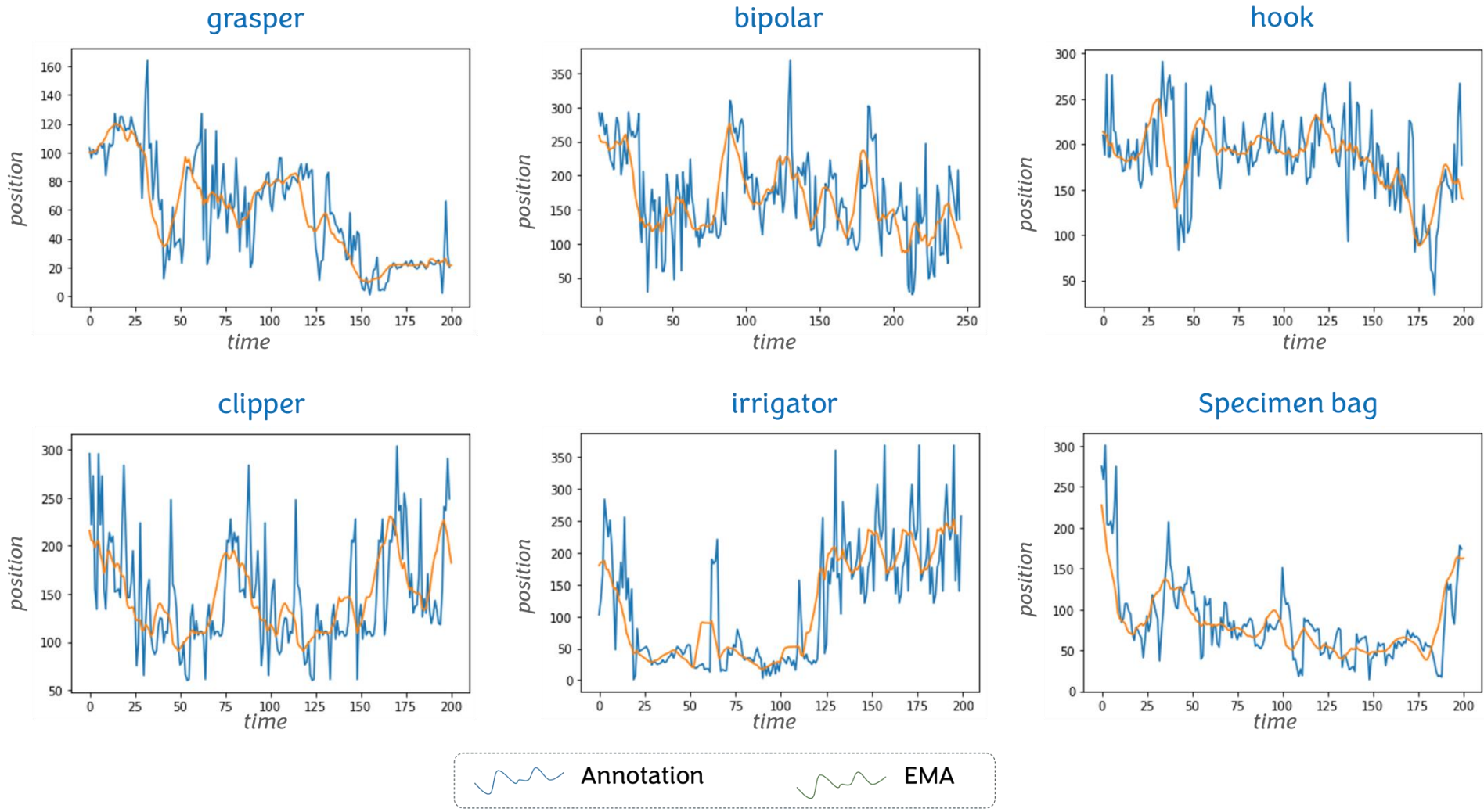
(d) Gallbladder dissection

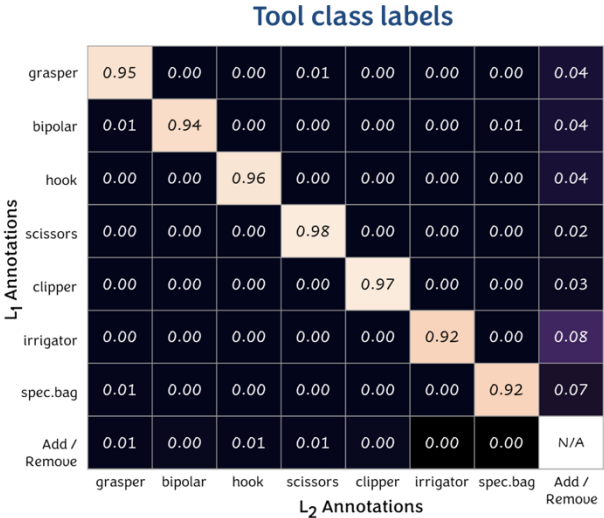


(e) Gallbladder packaging

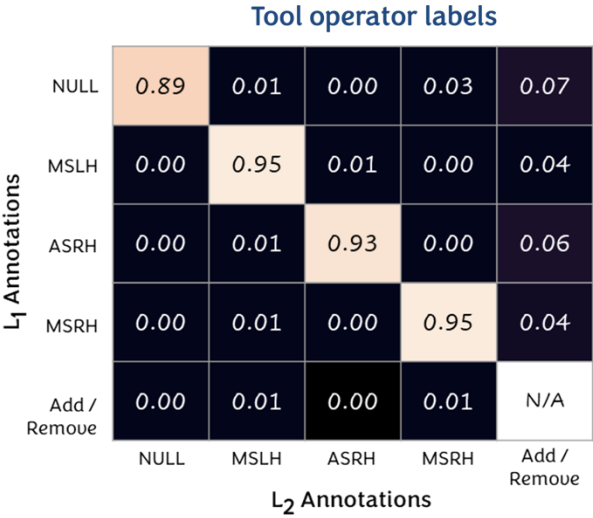


(f) Cleaning and coagulation

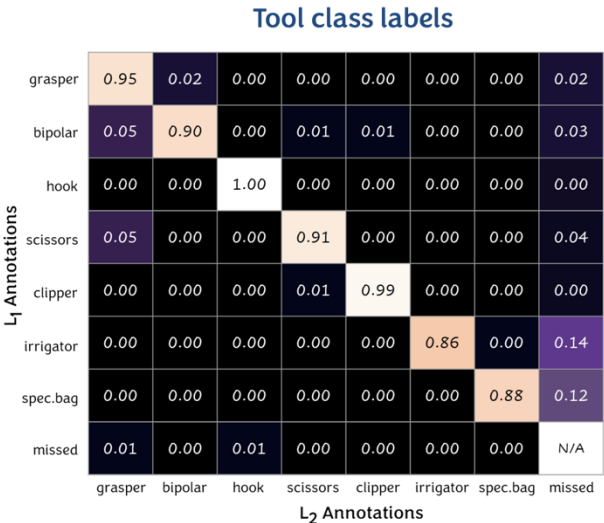




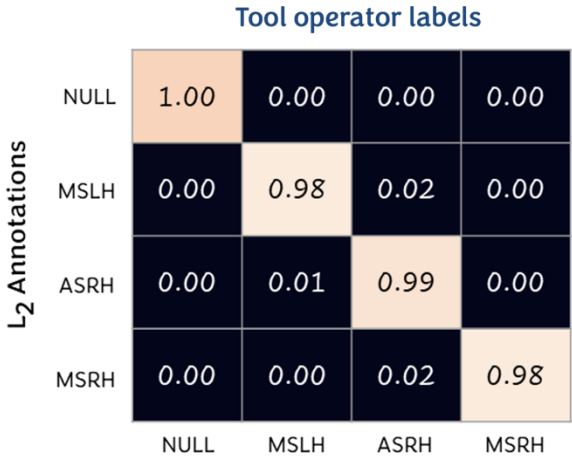
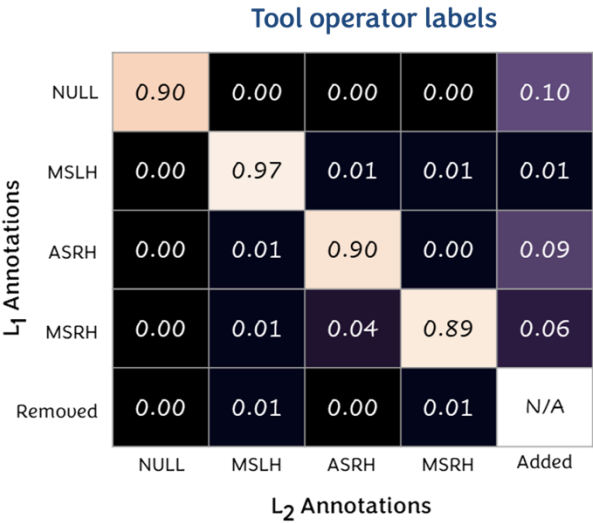
(a) Inter-rater agreement



	Intra-rater	Inter-rater
Bounding Box Jaccard Index	99.4	91.8
Tool Class Cohen's Kappa	94.6	95.2
Tool Operator Cohen's Kappa	94.0	92.7



(a) Inter-rater agreement



Mediated Labels
(c) Mediation agreement



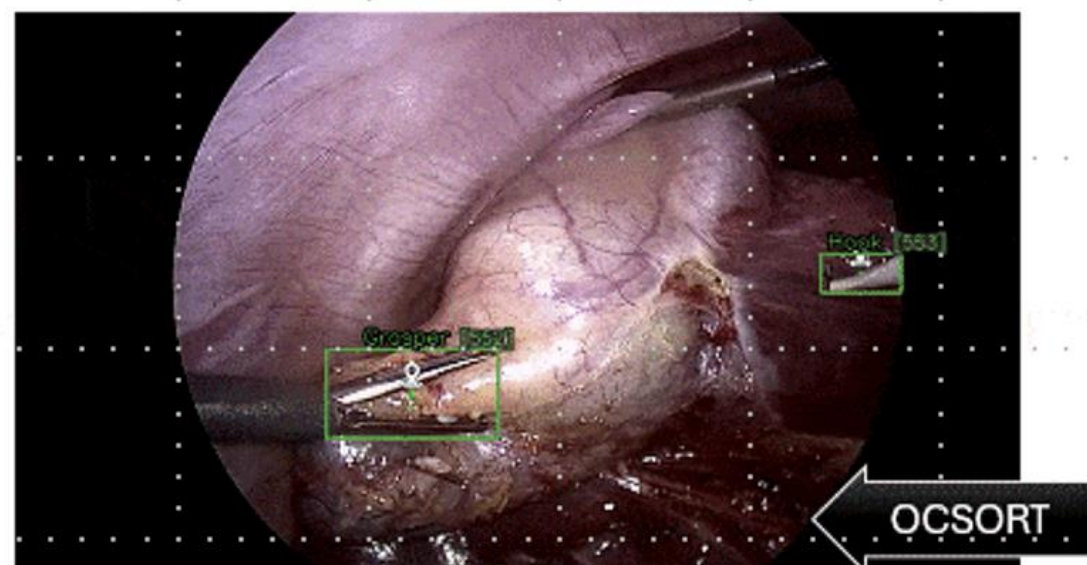
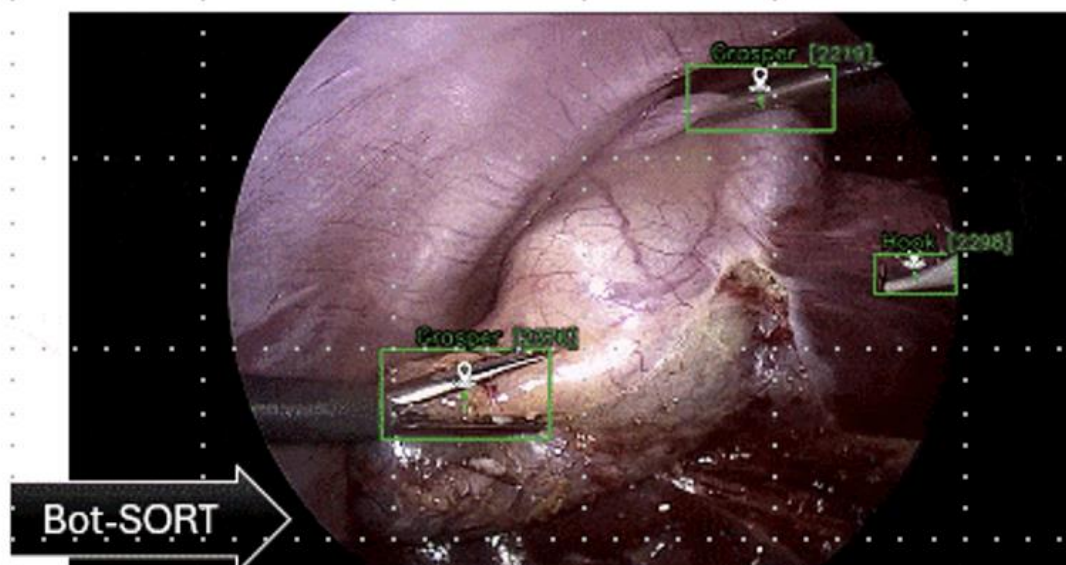
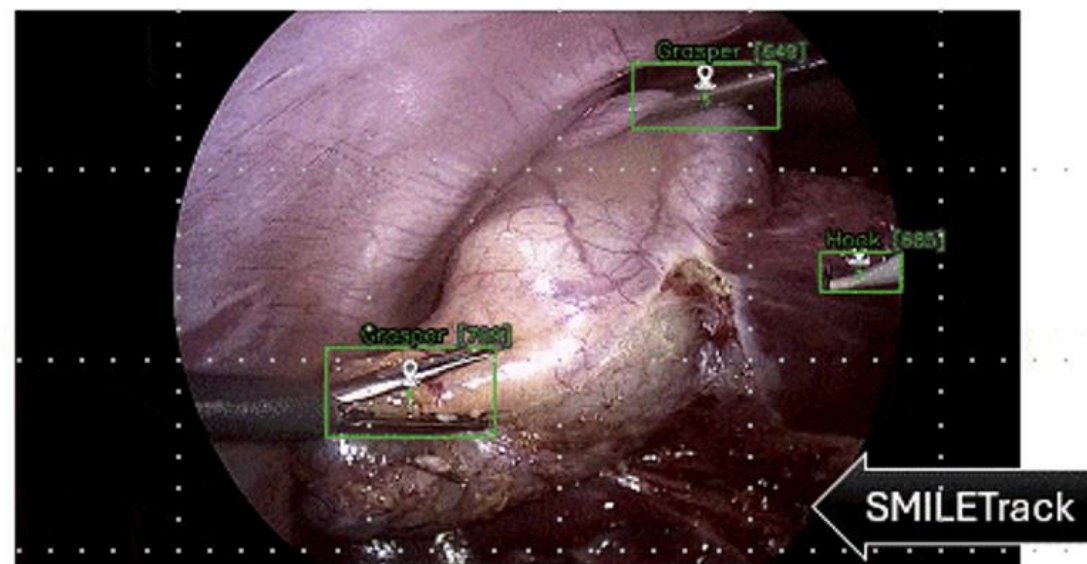
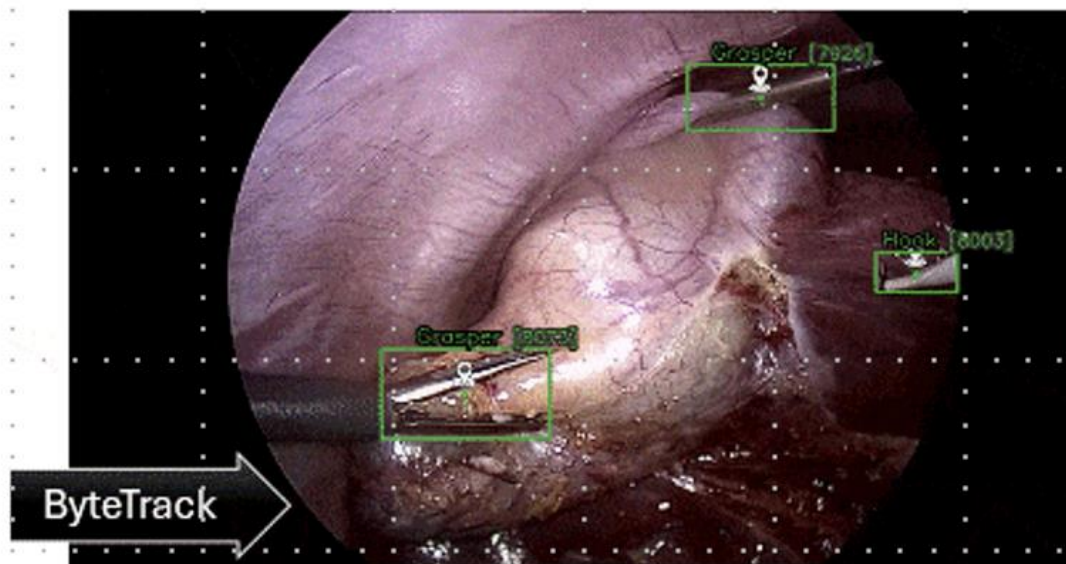
- 13 object detectors

Detector Model	Detection AP accross 3 thresholds			Detection AP per category. (% AP @ $\Theta = 0.5$)							Detection AP across surgical visual challenges								Speed	
	$AP_{0.5} \uparrow$	$AP_{0.75} \uparrow$	$AP_{0.5:0.95} \uparrow$	Grasper	Bipolar	Hook	Scissors	Clipper	Irrigator	Bag	Bleeding	Blur	Smoke	Crowded	Occluded	Reflection	Foul	Lens	Trocar	FPS \uparrow
Faster-RCNN [43]	56.0	38.1	34.6	53.5	65.0	80.1	60.9	70.1	26.8	31.8	57.9	41.0	54.5	43.5	55.0	46.9	41.2	35.7		7.6
Cascade-RCNN [7]	51.7	39.0	34.7	52.0	58.9	79.7	45.7	44.9	23.7	17.9	53.9	39.0	48.1	39.5	46.4	29.1	33.7	33.7		7.0
CenterNet [69]	53.0	39.5	35.0	60.2	61.4	86.4	56.3	68.0	25.8	10.2	58.0	42.1	50.2	36.7	51.7	46.0	35.8	30.8		33.8
FCOS [54]	43.5	31.5	28.1	51.2	44.3	74.7	49.2	54.2	21.9	7.2	47.8	40.6	51.5	15.1	40.8	42.7	29.7	17.6		7.7
SSD [30]	61.9	37.8	36.1	75.2	62.2	91.6	63.4	72.9	22.5	40.8	64.5	49.3	58.3	57.5	62.4	53.9	47.7	42.6		30.9
PAA [23]	64.5	44.9	41.1	69.6	79.0	89.2	68.7	74.2	37.6	28.9	67.1	55.6	65.0	55.0	64.6	56.0	51.2	47.5		8.5
Def-DETR [71]	58.4	42.0	38.3	60.6	66.5	83.8	61.9	72.0	39.9	23.8	62.4	42.7	58.6	37.1	57.4	43.9	41.5	47.4		10.2
Swin-T [31]	62.3	44.3	40.2	63.3	64.8	83.0	80.2	77.2	38.0	26.8	63.5	53.8	62.8	35.3	61.1	66.2	55.2	45.7		9.8
YOLOX [19]	64.7	48.9	44.2	69.6	72.2	89.4	75.4	79.1	37.3	27.1	68.2	55.6	66.0	45.9	64.2	52.5	58.1	43.1		23.6
YOLOv7 [62]	80.6	62.0	56.1	90.5	86.4	96.0	82.3	89.3	49.1	66.2	80.2	61.2	80.1	79.5	82.1	65.6	71.2	66.7		20.6
YOLOv8 [58]	79.1	62.4	55.6	87.9	84.5	96.2	80.0	87.2	48.4	65.0	77.1	58.3	74.4	76.2	80.4	70.3	57.4	62.9		29.0
YOLOv9 [61]	80.2	62.6	56.5	88.5	87.6	96.0	79.3	87.1	50.1	67.7	78.1	54.0	78.2	78.6	81.1	65.3	63.4	63.1		23.7
YOLOv10 [60]	80.1	62.1	55.8	87.6	86.6	96.0	81.9	89.0	53.8	61.3	77.8	61.9	78.7	77.5	81.2	66.7	59.3	65.4		28.6

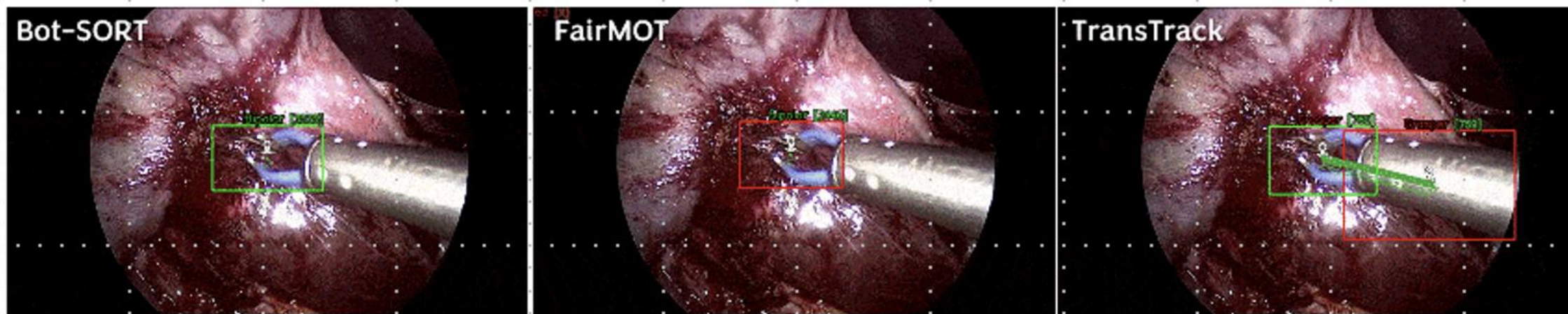
- 8 object trackers

Model	HOTA Metrics				CLEAR Metrics					Identity Metrics			Count Metrics		Speed
Tracker	HOTA↑	DetA↑	LocA↑	AssA↑	MOTA↑	MOTP↑	MT↑	PT↓	ML↓	IDF1↑	IDSW↓	Frag↓	#Dets	#IDs	FPS↑
<i>Intraoperative Trajectory (Groundtruth counts: #Dets = 29994, #IDs = 70)</i>															
OCSORT [33]	14.6	52.7	86.7	4.1	49.2	85.0	24	32	14	9.5	2921	2731	21936	3336	10.2
FairMOT [66]	5.8	25.8	75.9	1.3	5.0	73.9	3	24	43	4.3	4227	1924	15252	4456	14.2
TransTrack [50]	7.4	31.5	84.4	1.7	4.2	82.9	9	36	25	4.2	4757	1899	21640	4079	6.7
ByteTrack [67]	15.8	70.6	85.7	3.6	67.0	84.0	54	12	2	9.5	4648	2429	28440	5383	16.4
Bot-SORT [1]	17.4	70.7	85.4	4.4	69.6	83.7	58	11	1	10.2	3907	2376	29302	4501	8.7
SMILETrack [63]	15.9	71.0	85.5	3.7	66.4	83.8	55	13	2	9.2	4968	2369	28821	5761	11.2
<i>Intracorporeal Trajectory (Groundtruth counts: #Dets = 29994, #IDs = 247)</i>															
OCSORT [33]	23.7	51.4	86.5	11.0	47.1	84.8	115	87	45	18.1	2953	2796	21797	3526	10.2
FairMOT [66]	7.5	19.7	76.1	2.9	5.4	74.0	19	60	168	6.0	2890	1496	11287	3962	14.2
TransTrack [50]	13.1	31.5	84.4	5.5	4.6	82.9	80	79	88	8.7	4648	1791	21640	4079	6.7
ByteTrack [67]	24.7	70.6	85.7	8.7	67.4	84.0	176	48	23	16.9	4515	2290	28440	5383	16.4
Bot-SORT [1]	27.0	70.7	85.4	10.4	70.0	83.7	188	38	21	18.9	3771	2238	29300	4501	8.7
SMILETrack [63]	24.9	66.7	85.5	8.9	66.7	83.8	186	39	22	16.9	4868	2232	28820	5779	11.2
<i>Visibility Trajectory (Groundtruth counts: #Dets = 29994, #IDs = 916)</i>															
SORT [4]	17.4	39.5	85.2	7.8	21.4	83.3	139	399	378	13.4	6619	2138	16595	8844	19.5
OCSORT [33]	37.0	52.6	86.5	26.2	50.2	84.8	300	371	245	35.9	2317	2260	22197	3587	10.2
FairMOT [66]	15.3	25.0	75.8	9.5	7.1	73.7	58	218	640	14.4	3140	1574	15338	4875	14.2
TransTrack [50]	19.2	31.6	84.4	11.8	5.8	82.9	224	280	412	16.1	4273	1403	21640	4079	6.7
ByteTrack [67]	41.5	70.7	85.7	24.8	69.3	84.0	591	217	108	36.8	3930	1704	28440	5383	16.4
Bot-SORT [1]	44.7	70.8	85.5	28.7	72.0	83.7	638	184	94	41.4	3183	1638	29300	4505	8.7
SMILETrack [63]	41.3	71.0	85.6	24.4	68.9	83.8	619	192	105	36.5	4227	1641	28821	5752	11.2

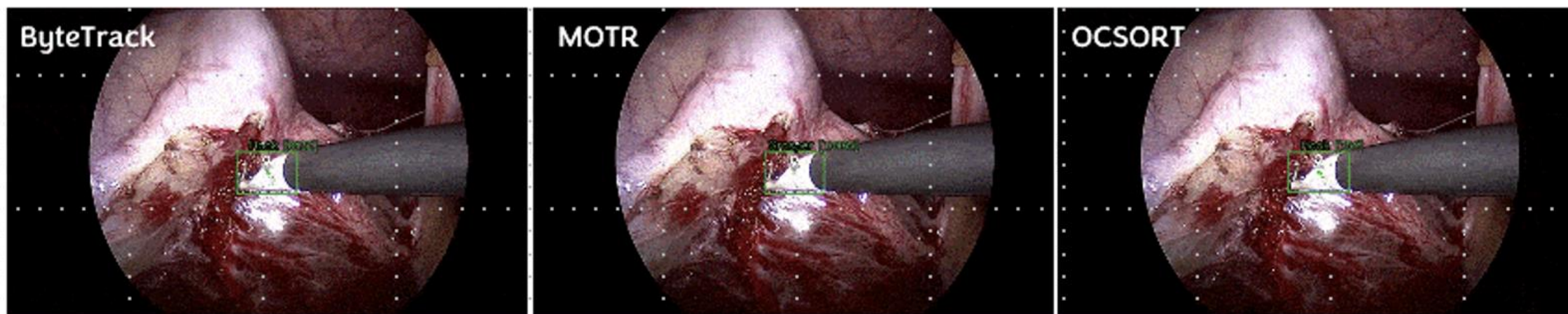
Qualitative Results of SOTA Tracking Models



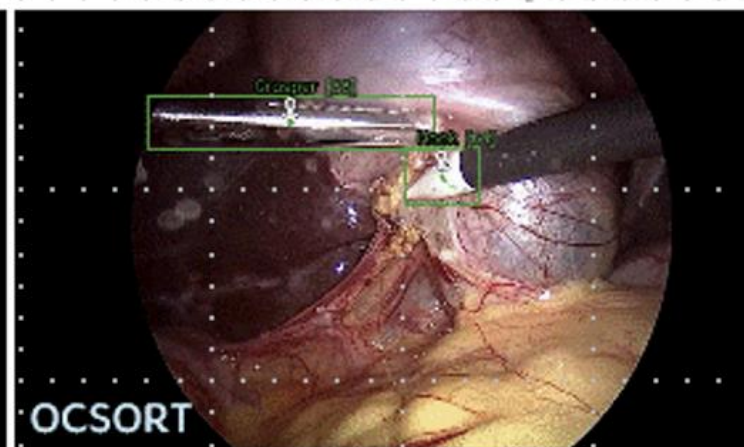
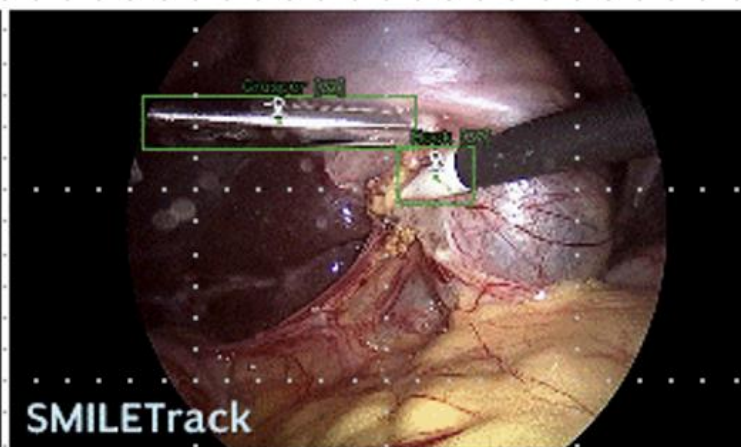
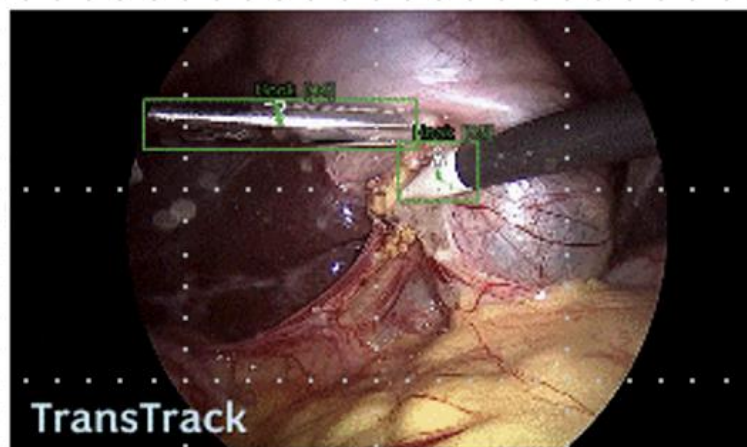
Tracking Performance Across Surgical Visual Challenges



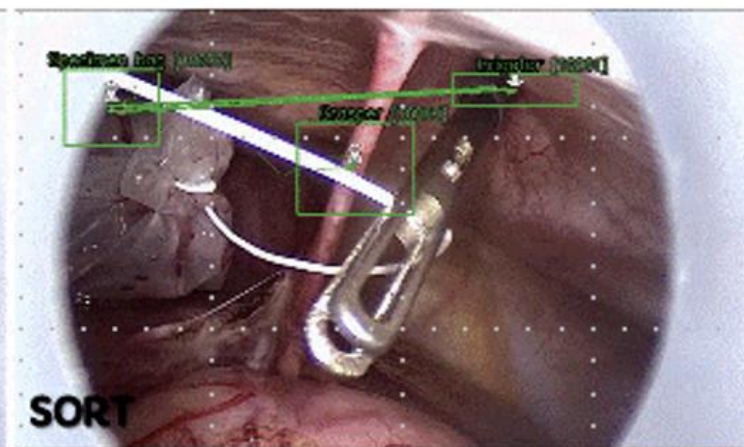
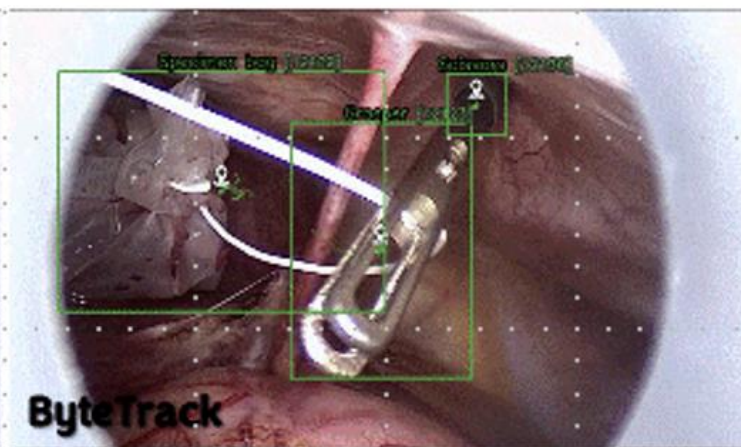
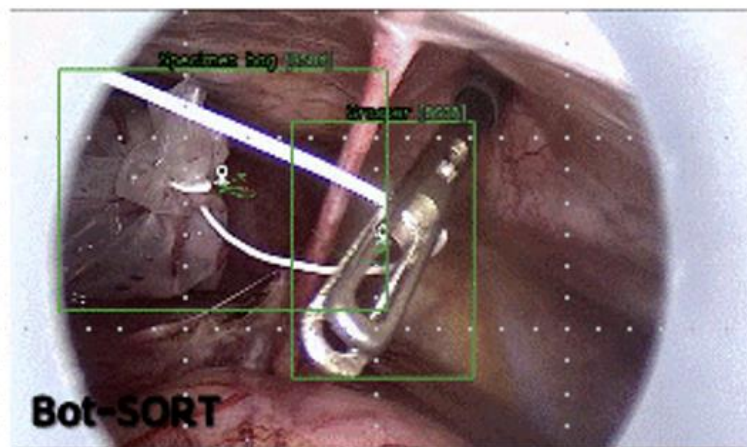
Tracking in Smoking Scene



Tracking in Bleeding Scene

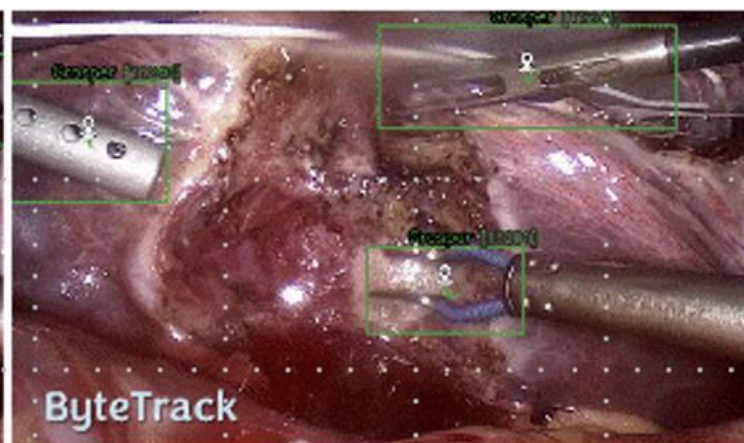
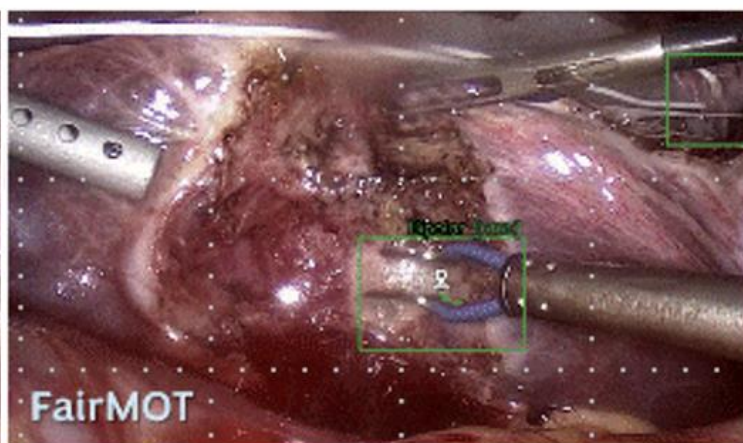
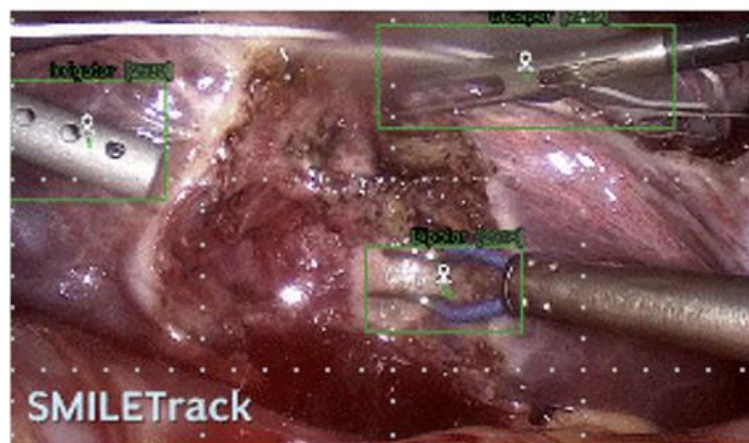


Tracking under Occlusion

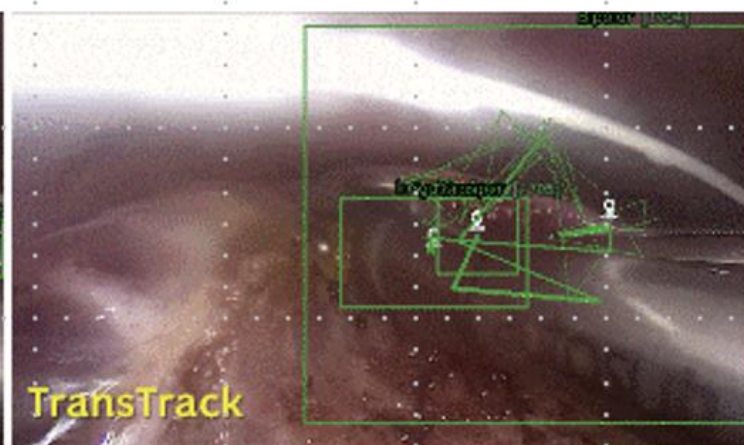
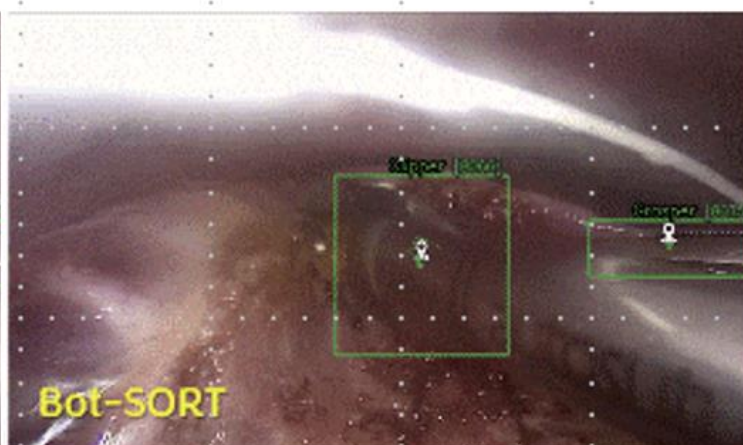
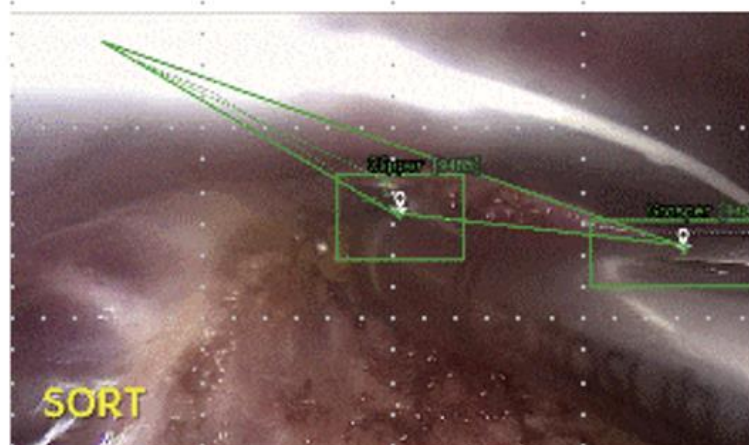


Tracking in Trocar View

Tracking Performance Across Surgical Visual Challenges

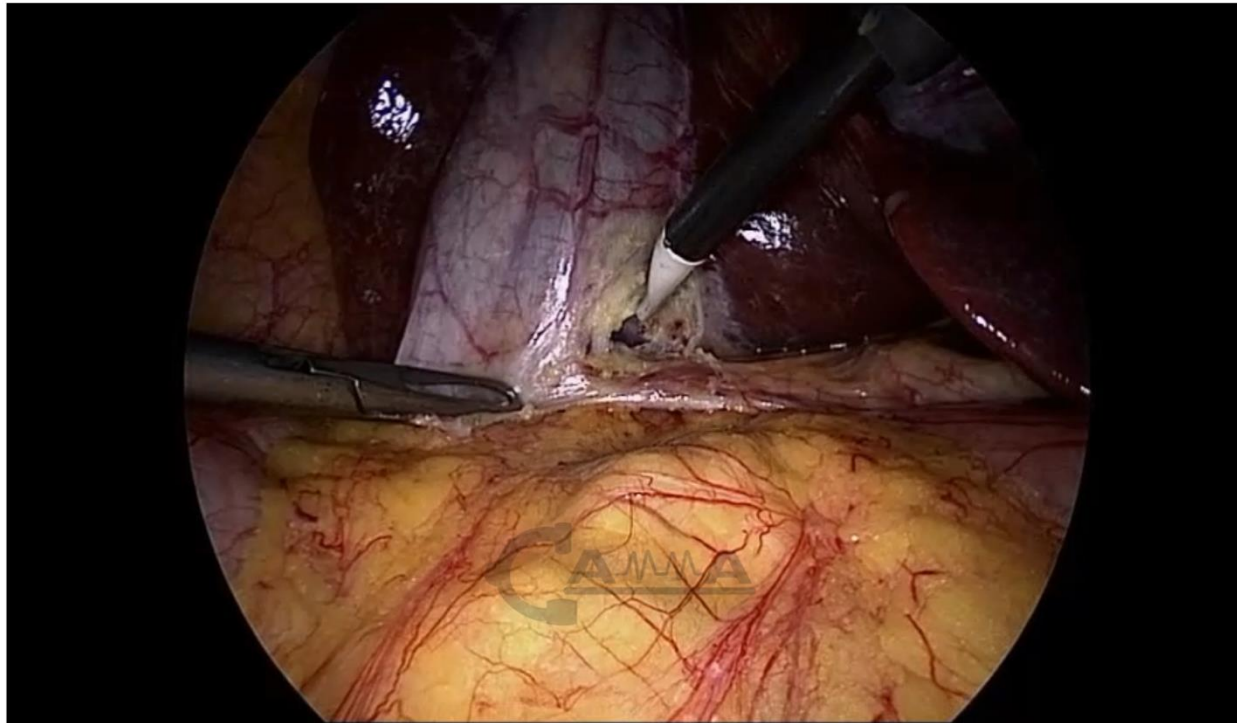


Tracking in Waterlogged Scene



Tracking under Specular Light Reflection

Analyzing Tracking Performance in 2D



SORT

BotSORT

OCSORT

SMILETrack

FairMOT

MOTR

ByteTrack

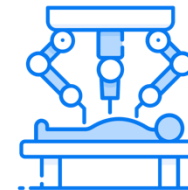
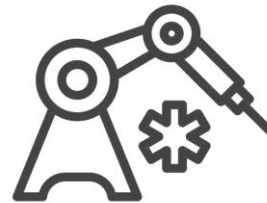
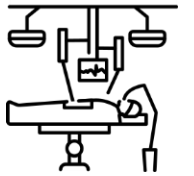
TransTrack

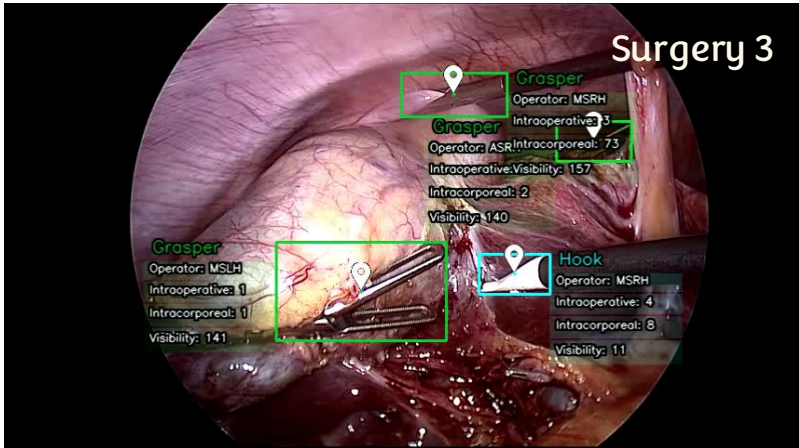
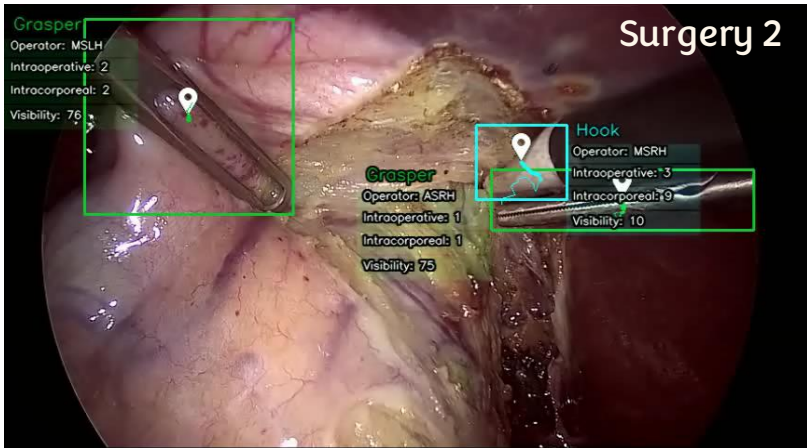
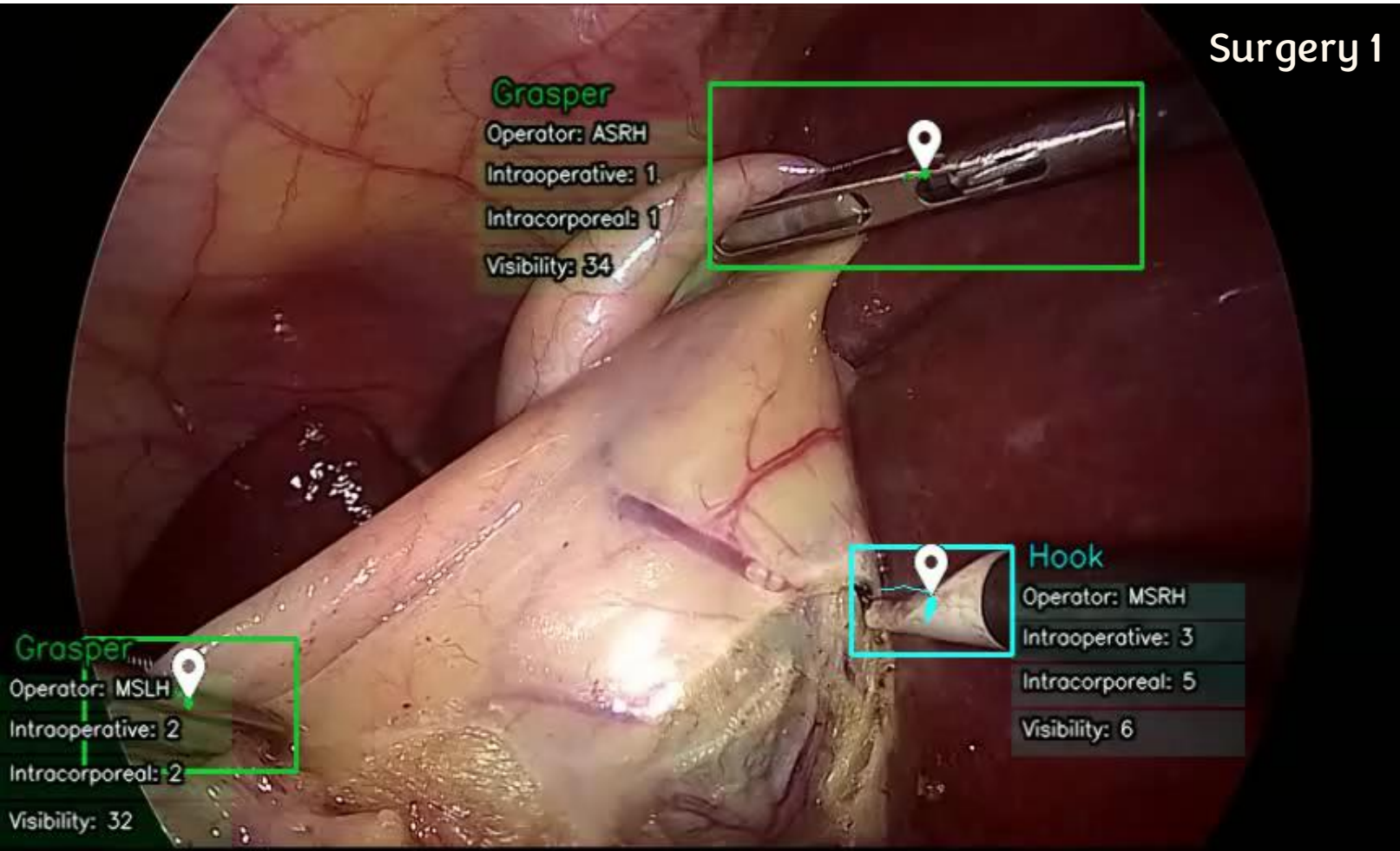
Tool icon = predicted tool class and location, tracking @ 25 FPS | **Transparent tool phantom = ground truth @ 1 FPS**
Coloured lines = tool trajectories | **Change in trajectory colour = identity switch** | **Broken lines = track fragmentation**

Facilitates the development of robust AI systems for surgical assistance.



With better potential applications in skill assessment, safety zone estimation, navigation, and enhancing human-machine collaboration in the operating room.





► SurgiTrack jointly track all the 3 perspectives . . .

- Multi-camera pedestrian & cross-view tracking
- Cross-view learning from autonomous driving
- Egocentric & robot-assisted video analysis
- Task-aware tracking in robotic surgery
- Integrating depth & kinematics for richer insight
- Benchmarking next-gen surgical tracking models

CholecTrack20

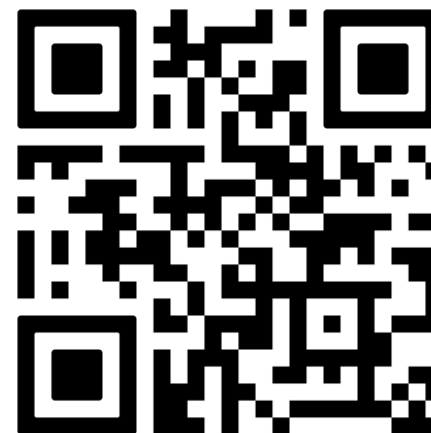
A Multi-Perspective Tracking Dataset for Surgical Tools



Paper



Dataset



Code