A66 (Wed-PM)

# LiteTracker
## Leveraging Temporal Causality for Accurate Low-latency Tissue Tracking

Mert Asım Karaoğlu, Wenbo Ji, Ahmed Abbas, Nassir Navab, Benjamin Busam, and Alexander Ladikos

ImFusion

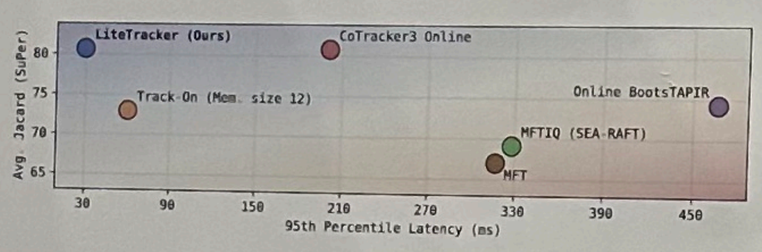
CAMP TUM

## Motivation

Accurate tissue tracking is crucial for **surgical navigation** and **XR** applications, yet remains difficult due to non-rigid deformations, occlusions, and abrupt camera motions.

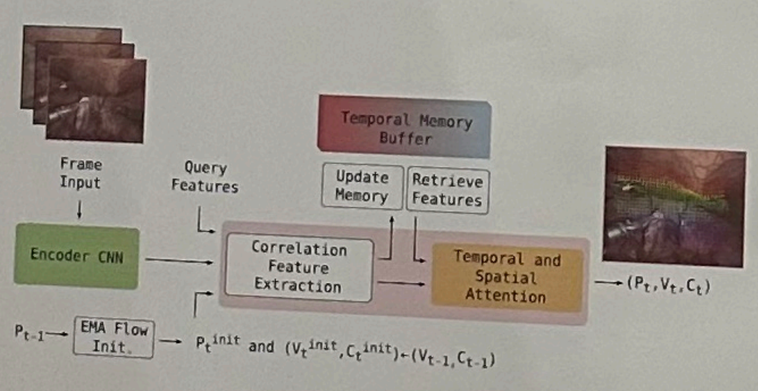
While existing approaches that formulate the task as long-term point tracking achieve high accuracy, their high latency hinders real-time use.

**LiteTracker** builds on the prior art and bridges this gap by introducing a set of **training-free** runtime optimizations that enable **efficient** and **accurate frame-by-frame** tissue tracking.




## Overview

LiteTracker tracks a set of query points by estimating their motion ($P_t$), visibility ($C_t$), and confidence score ($V_t$).




It uses a lightweight CNN to extract point descriptors and employs a **temporal memory buffer** that caches point-to-template correlation features to efficiently retain causal information.

For context fusion, it leverages consecutive temporal and spatial attention blocks, employing **per-point causal attention masks to** enable **dynamic removal or addition of query points** at runtime.

Each frame's motion prediction is warm-started with an **exponential moving average (EMA) flow** for fast convergence.

Built on the CoTracker3[1] architecture, LiteTracker remains fully compatible with its pre-trained weights.

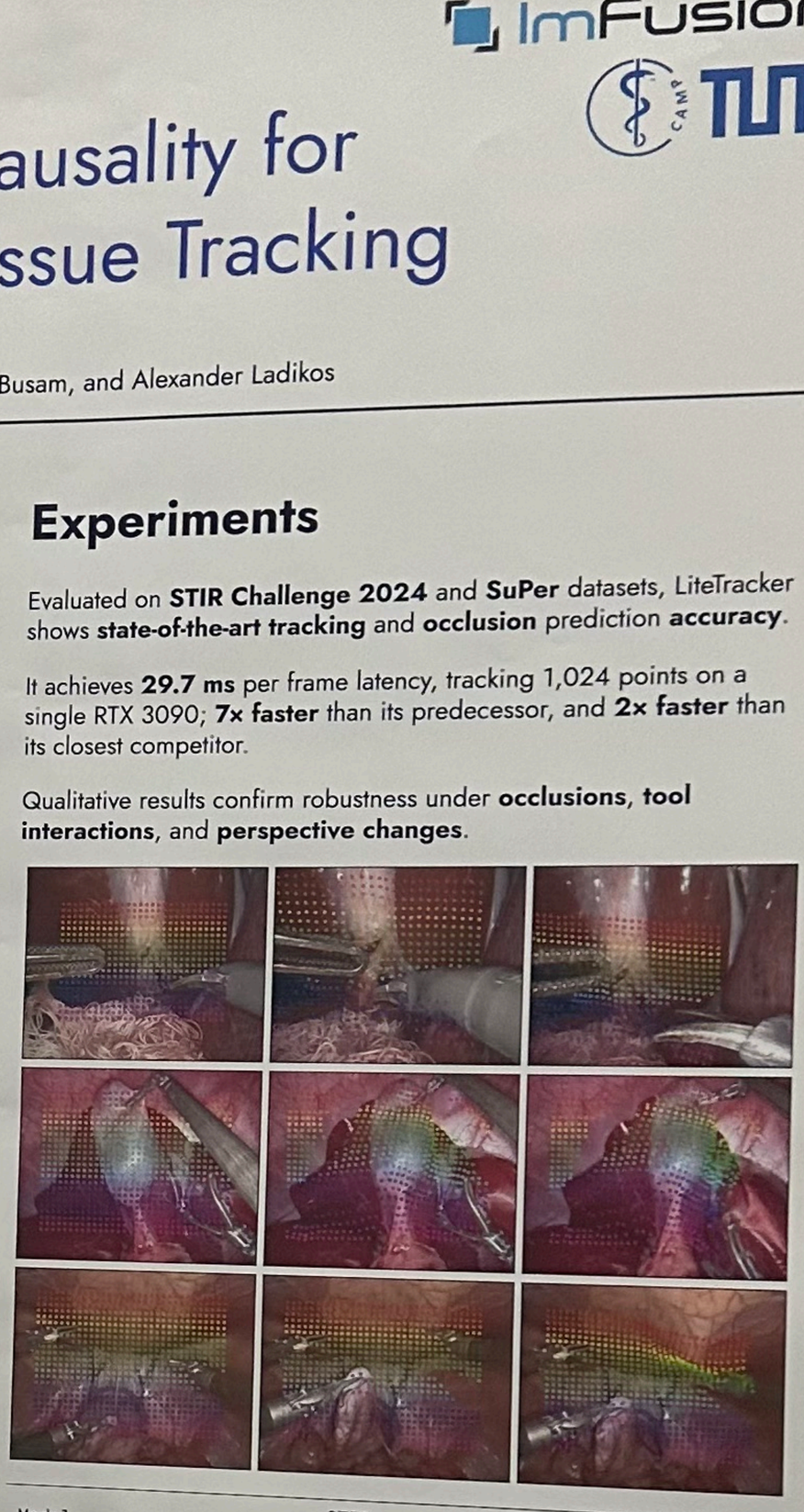
*Live demos at our booth!*

## Experiments

Evaluated on **STIR Challenge 2024** and **SuPer** datasets, LiteTracker shows **state-of-the-art tracking** and **occlusion** prediction **accuracy**.

It achieves **29.7 ms** per frame latency, tracking 1,024 points on a single RTX 3090; **7× faster** than its predecessor, and **2× faster** than its closest competitor.

Qualitative results confirm robustness under **occlusions**, **tool interactions**, and **perspective changes**.



| Model | Input | STIR | SuPer | | Latency (ms) ↓ |
|---|---|---|---|---|---|
| | | $\delta^{avg}$ ↑ | AJ ↑ | OA ↑ | |
| A-MFST* | Frame | 58.59 | -- | -- | -- |
| MFTIQ (ROMA)* | Frame | 77.22 | 66.61 | 83.73 | 4355.11 |
| Online BootsTAPIR | Frame | 68.59 | 74.27 | 93.63 | 466.38 |
| MFTIQ (SEA-RAFT)* | Frame | 76.82 | 69.14 | 83.73 | 327.93 |
| MFT* | Frame | **77.62** | 67.02 | 83.73 | 317.05 |
| CoTracker3 Online | Win. | 75.24 | **80.82** | 96.96 | 200.98 |
| Track-On (48) | Frame | 74.44 | 70.22 | 83.92 | 74.80 |
| Track-On (12) | Frame | 72.74 | 72.97 | 86.86 | 60.18 |
| **LiteTracker (Ours)** | Frame | 75.81 | 80.68 | **97.45** | **29.67** |

## Conclusion

LiteTracker is a **low-latency** and **accurate tissue tracking method** for real-time surgical applications.

It combines **efficiency** and **robustness**, achieving substantial speedups while preserving accuracy.

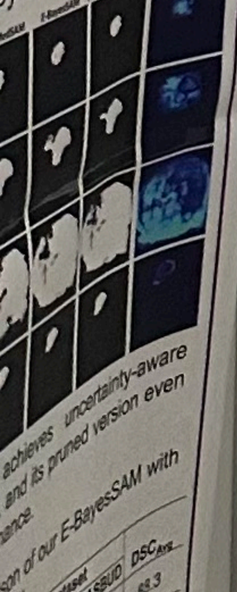
Its design enables seamless integration into **real-time surgical robotics** and **XR** applications.

1- Karaev, Nikita, Iurii Makarov, Jianyuan Wang, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. "Cotracker3: Simpler and better point tracking by pseudo-labelling real videos." arXiv preprint arXiv:2410.11831 (2024).

ong Key Laboratory of
nen University Medical
Shenzhen Maternity and
School of Biomedical
epartment of Biomedical
Cleveland, OH, USA.

ENTAL SETUP
SA-Med2D-20M
ive publicly available
sets——DDTI, TN3K,
OASBUD.
arity Coefficient (DSC).
AM (LiteMedSAM)
X 3070 (8GB VRAM)
W (initial LR 0.0001)

ENTAL RESULTS
ccessfully adds reliable
ation while maintaining
able to the MedSAM
% $DSC_{Avg}$). SO-KAN
non-essential tokens,
y 0.7%.
peed of 0.03s/sample,
tly outperforms other
n methods, making it
oyment.



achieves uncertainty-aware
and its pruned version even
ance.

son of our E-BayesSAM with
(%).

| Dataset | BUSI | OASBUD | $DSC_{Avg}$ |
|---|---|---|---|
| SAT | 90.6 | 80.9 | 88.3 |
| 1.0 | 90.5 | 80.8 | 88.1 |
| 6 | 90.5 | 80.8 | 88.0 |
| 6 | 90.2 | 84.7 | 89.0 |
| 8 | | | |

TA2021).