SAM-EM: Real-Time Segmentation for Automated Liquid Phase Transmission Electron Microscopy

Alexander Wang [*]
Georgia Institute of Technology
awang652@gatech.edu

Max Xu [*]
Northwestern University
maxxu@u.northwestern.edu

Risha Goel
Georgia Institute of Technology
rgoel63@gatech.edu

Zain Shabeeb
Georgia Institute of Technology
zshabeeb3@gatech.edu

Isabel Panicker
Georgia Institute of Technology
ipanicker@gatech.edu

Vida Jamali [†]
Georgia Institute of Technology
vida@gatech.edu

Abstract

The absence of robust segmentation frameworks for noisy liquid phase transmission electron microscopy (LPTEM) videos prevents reliable extraction of particle trajectories, creating a major barrier to quantitative analysis and to connecting observed dynamics with materials characterization and design. To address this challenge, we present Segment Anything Model for Electron Microscopy (SAM-EM), a domain-adapted foundation model that unifies segmentation, tracking, and statistical analysis for LPTEM data. Built on Segment Anything Model 2 (SAM 2), SAM-EM is derived through full-model fine-tuning on 46,600 curated LPTEM synthetic video frames, substantially improving mask quality and temporal identity stability compared to zero-shot SAM 2 and existing baselines. Beyond segmentation, SAM-EM integrates particle tracking with statistical tools, including mean-squared displacement and particle displacement distribution analysis, providing an end-to-end framework for extracting and interpreting nanoscale dynamics. Crucially, full fine-tuning allows SAM-EM to remain robust under low signal-to-noise conditions, such as those caused by increased liquid sample thickness in LPTEM experiments. By establishing a reliable analysis pipeline, SAM-EM transforms LPTEM into a quantitative single-particle tracking platform and accelerates its integration into data-driven materials discovery and design. Project page: github.com/JamaliLab/SAM-EM.

1 Introduction

In situ liquid phase transmission electron microscopy (LPTEM) is an emerging microscopy technique with an unparalleled combination of spatial and temporal resolution, enabling direct visualization of the nanoscale dynamics of materials in liquid solutions using an electron microscope [18, 9, 23, 10]. These features make LPTEM a powerful tool for single-particle tracking and for probing how nanoparticles nucleate, self-assemble, and transform under native conditions [6, 12, 3, 8, 11]. Looking

---

[*]Equal contribution.
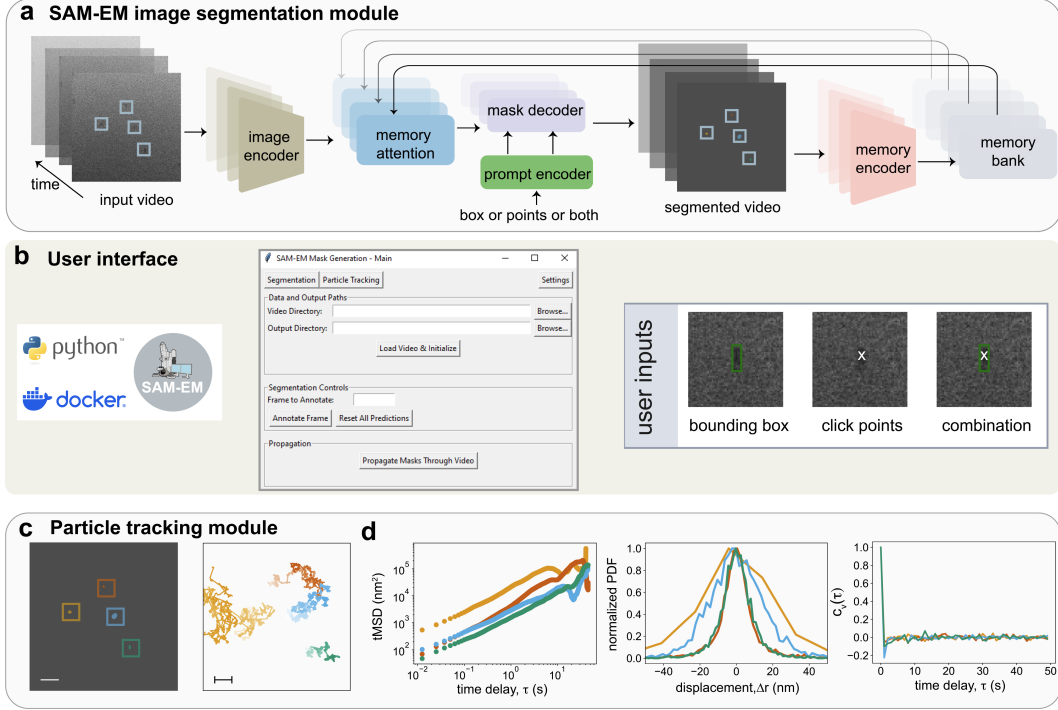
[†]Corresponding author. vida@gatech.edu

Figure 1: **SAM-EM framework for particle segmentation and tracking in LPTEM. a**, The segmentation module segments an LPTEM video with multiple nanoparticles, each prompted once on the first frame. **b**, Deployment on any Python or Docker system with a graphical interface that enables particle selection via bounding-box prompts and statistical test selection. **c**, Extracted trajectories undergo quantitative analyses, including spatiotemporal paths (color-coded over 0-45 seconds), mean squared displacements, displacement distributions, and velocity autocorrelations. Scale bar, 250 nm.

ahead, the role of in situ electron microscopy is poised to evolve from a passive characterization technique into an active feedback-driven experimental platform [19, 7, 17, 2]. In such a paradigm, the microscope becomes not only an imaging instrument but also a controllable environment, where experimental knobs such as electron beam dose rate, ionic concentration, or fluid mixing can be modulated in real time. Coupled with on-the-fly analysis, this feedback allows researchers to actively steer nanoscale assembly and transformation, turning LPTEM into a platform for rapid hypothesis testing and iterative exploration, therefore accelerating materials discovery. Realizing this vision requires analysis pipelines that operate during acquisition, extracting particle trajectories in real time, and feeding them back to the microscope to guide automated control of experimental parameters.

Recent advances in vision foundation models suggest a path toward this goal. The Segment Anything Model 2 (SAM 2), a vision transformer trained on 50.9K videos with 93.7 million parameters, was released by Meta in 2024 as a generalization of its predecessor SAM. SAM 2 supports promptable video segmentation at real-time speeds of nearly 50 frames per second (fps), conditioning its predictions both on user prompts and on a memory bank of previously segmented frames. This design allows SAM 2 to identify and track objects that move in and out of frame or cross paths with one another, features that are critical for particle tracking in LPTEM.

However, when applied directly (zero-shot) to noisy LPTEM videos, SAM 2 fails to achieve robust performance, particularly under low signal-to-noise ratio (SNR) conditions caused by thick liquid layers. Image quality deteriorates with increasing liquid thickness due to inelastic electron scattering. SAM 2's performance is further challenged by the stochastic dynamics of the nanoparticles, including large frame-to-frame displacements and transient departures from the field of view. Existing deep learning approaches for static TEM segmentation [24, 5, 21, 7, 22, 4, 15, 25] likewise generalize poorly to LPTEM data and discard temporal context by processing frames independently.

2

To address these challenges, we introduce SAM-EM, a domain-adapted framework that extends SAM 2 through full model fine-tuning on 46,600 synthetic LPTEM video frames (Fig. 1). SAM-EM unifies segmentation, tracking, and statistical analysis into a single pipeline, integrating tools for mean-squared displacement (MSD), displacement distributions, and other relevant canonical statistical metric for spatiotemporal trajectories. Full fine-tuning enables SAM-EM to maintain segmentation fidelity under low-SNR conditions caused by thicker liquid samples, the most challenging yet also the most experimentally relevant regime for LPTEM. By combining robust video segmentation with quantitative analysis, SAM-EM moves LPTEM toward its envisioned role as a feedback-driven experimental platform for data-driven materials discovery.

2 Methods

2.1 Dataset Generation

To train and evaluate SAM-EM, we generated synthetic LPTEM videos with ground-truth masks following the method of Yao et al.[24]. The original code was modified to reflect experimental conditions (e.g., liquid thickness, particle size/shape, and electron beam dose rate). The particle positions were sampled from LEONARDO, a physics-informed variational autoencoder for stochastic diffusion in LPTEM [20]. The data was generated at 0.25 nm per pixel, corresponding to a 256 nm$\times$256 nm field of view (1024$\times$1024 pixels). By varying the simulated liquid thickness from 5 nm to 160 nm, we produced a dataset spanning a wide range of SNR values representative of experimental conditions (Appendix A). The final dataset consisted of 1,000 synthetic videos, each 50 frames in length, with ground-truth masks and centroid positions. We reserved 68 videos for validation and testing, and additionally generated longer test sequences from independently sampled LEONARDO trajectories to assess generalization beyond 50 frames. Example frames are shown in Appendix A.

2.2 SAM 2 Video Segmentation and Fine-tuning

We fine-tuned SAM 2 for LPTEM video segmentation following the protocol of Ravi et al. [16], using the synthetic datasets described in Section A. Of the 1,000 generated videos, 932 were used for training and the remainder reserved for validation and testing. Training examples included particle collisions and overlaps to expose the model to motion patterns characteristic of LPTEM data. Fine-tuning was performed using the official `train.py` script from the SAM 2 codebase, with the default model architecture and the training algorithm left unchanged. The primary configuration modification was to enforce box-prompt conditioning during training. Ground-truth masks were supplied as bounding box prompts, reflecting how users interact with the SAM-EM interface by drawing approximate boxes around particles. Consistent with the prior work in domain-adapted foundation models such as μSAM [1], we opted for full model fine-tuning rather than adapter-based methods, as it produced statistically superior $\mathcal{J}\&\mathcal{F}$ scores and qualitatively more stable masks and trajectories in our experiments. Additional details regarding our fine-tuning setup are provided in Appendix B.

2.3 U-Net LPTEM Image Segmentation Baseline

For benchmarking, we implemented the U-Net model from Yao et al. [24], which to our knowledge is the only deep learning–based segmentation model specifically developed for LPTEM data. We trained the model using the dataset provided by the authors for 20 epochs with a learning rate of 1×10^{-6}. Unlike SAM 2 and our fine-tuned SAM-EM framework, this U-Net model does not support prompt-based refinement, which limits its ability to adapt to complex or noisy scenarios. As a result, its segmentation performance was lower in our experiments. Nevertheless, U-Net has the advantage of efficiency; it produces predictions substantially faster and with lower computational cost compared to SAM 2 and SAM-EM.

3 Experiments

3.1 Experimental Setup

To evaluate the performance and robustness of SAM-EM, we constructed a testing set consisting of five distinct trials designed to capture a range of experimental conditions (details in Appendix A). The evaluation was performed on three standard segmentation and tracking metrics: Jaccard index (IoU), F1-score, and the composite $\mathcal{J}\&\mathcal{F}$ score. Formal definitions of these metrics are provided in Appendix A.1.
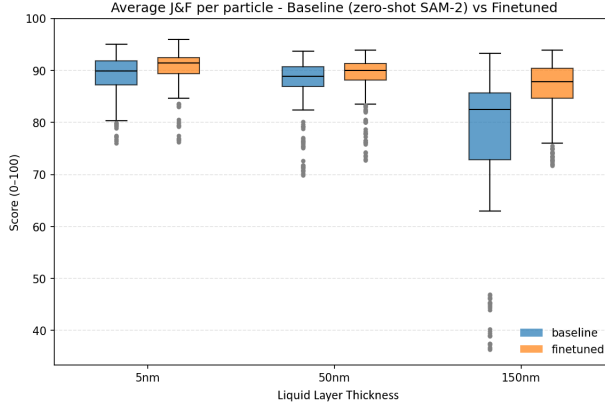
3.2 Results

Figure 2: **Performance accuracy of SAM-EM in segmentation on simulated datasets.** Comparison of fine-tuned SAM-EM model and baseline (zero-shot SAM 2) on a simulated test dataset with 300 frames with two moving nanoparticles and with minimal particle interaction.

Fig. 2 shows SAM-EM consistently outperforms the zero-shot SAM 2 baseline on LPTEM segmentation as measured by the $\mathcal{J}\&\mathcal{F}$ score (higher, better). Improvements are modest for thin liquid layers (5 nm and 50 nm) but become pronounced under thicker liquid conditions, where fine-tuning yields a significant performance gain. This regime is particularly important because most commercial LPTEM chambers employ lithographically defined spacers of either 50 nm or 150 nm; the latter is frequently used as it imposes less confinement on the specimen. However, increased liquid thickness leads to additional electron scattering, lowering the image contrast and effective SNR, which makes segmentation more challenging. Strong performance in this regime therefore demonstrates both practical relevance and technical robustness.

The U-Net baseline achieved median $\mathcal{J}\&\mathcal{F}$ scores of approximately 0.5 across thicknesses, confirming that SAM 2 and SAM-EM substantially outperform existing LPTEM-specific segmentation approaches under the evaluated conditions.

Following segmentation, trajectory analysis on the extracted masks demonstrated reliable multi-particle tracking (Appendix D). The reconstructed particle paths are physically realistic, with minimal identity switching observed except under highly complex video conditions, where the video contains an unusually large number of particles that overlap in many of the frames. These results support the applicability of SAM-EM for tracking nanoparticle dynamics during self-assembly processes in LPTEM experiments.

4 Conclusion

We presented SAM-EM, a domain-adapted foundation model that unifies segmentation, tracking, and statistical analysis for LPTEM videos. Through full model fine-tuning of SAM 2 on curated synthetic datasets, SAM-EM achieved significant improvements in mask fidelity and temporal identity stability, with particularly strong gains under low SNR conditions caused by increased liquid sample thickness. Beyond segmentation accuracy, SAM-EM enables interactive selection of particles and

immediate visualization of trajectory statistics such as mean-squared displacement and displacement distributions. This integration of robust segmentation with interpretable analysis opens a path toward closed-loop LPTEM workflows, where on-the-fly tracking can support automated experimentation. Such capabilities could ultimately allow researchers to probe defect dynamics in nanocrystals or follow the self-assembly of nanoparticles into ordered superlattices, directly connecting imaging to materials characterization and design. While our main evaluations use physics-grounded synthetic datasets, we also include qualitative results on real experimental videos and centroid-agreement analysis under noisy human-based annotation (Appendix E).

Although SAM-EM demonstrates strong performance, there are natural areas for further development. For example, distinguishing particles during severe overlap remains challenging, and model performance may vary across experimental setups due to differences in microscope conditions. However, these limitations are not fundamental. They can be addressed by expanding training datasets to include multi-lab experimental data. We view SAM-EM as an important first step toward a more generalizable framework and anticipate that community-driven data sharing and iterative refinement will further strengthen its utility across diverse in situ electron microscopy studies.

5 Acknowledgment

References

[1] Anwai Archit, Luca Freckmann, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Carolin Teuber, Melanie Spitzner, Constanza Tapia Contreras, Genevieve Buckley, Sebastian von Haaren, Sagnik Gupta, Marian Grade, Matthias Wirth, Günter Schneider, Andreas Dengel, Sheraz Ahmed, and Constantin Pape. Segment anything for microscopy. *Nature Methods*, 22(3):579–591, Mar 2025. ISSN 1548-7105.

[2] Marc Botifoll, Ivan Pinto-Huguet, and Jordi Arbiol. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. *Nanoscale horizons*, 7(12):1427–1477, 2022.

[3] Qian Chen, Jong Min Yuk, Matthew R Hauwiller, Jungjae Park, Kyun Seong Dae, Jae Sung Kim, and A Paul Alivisatos. Nucleation, growth, and superlattice formation of nanocrystals observed in liquid cell transmission electron microscopy. *MRS Bulletin*, 45(9):713–726, 2020.

[4] Bin Cheng, Enze Ye, He Sun, and Huan Wang. Deep learning-assisted analysis of single molecule dynamics from liquid-phase electron microscopy. *Chemical Communications*, 59(12): 1701–1704, 2023.

[5] James P Horwath, Dmitri N Zakharov, Rémi Mégret, and Eric A Stach. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *npj Computational Materials*, 6(1):108, 2020.

[6] Vida Jamali, Cory Hargus, Assaf Ben-Moshe, Amirali Aghazadeh, Hyun Dong Ha, Kranthi K Mandadapu, and A Paul Alivisatos. Anomalous nanoparticle surface diffusion in lctem is revealed by deep learning-assisted analysis. *Proceedings of the National Academy of Sciences*, 118(10):e2017616118, 2021.

[7] Sergei V Kalinin, Debangshu Mukherjee, Kevin Roccapriore, Benjamin J Blaiszik, Ayana Ghosh, Maxim A Ziatdinov, Anees Al-Najjar, Christina Doty, Sarah Akers, Nageswara S Rao, et al. Machine learning for automated experimentation in scanning transmission electron microscopy. *npj Computational Materials*, 9(1):227, 2023.

[8] Binbin Luo, John W Smith, Zihao Ou, and Qian Chen. Quantifying the self-assembly behavior of anisotropic nanoparticles using liquid-phase transmission electron microscopy. *Accounts of chemical research*, 50(5):1125–1133, 2017.

[9] Utkur Mirsaidov, Joseph P Patterson, and Haimei Zheng. Liquid phase transmission electron microscopy for imaging of nanoscale processes in solution. *MRS Bulletin*, 45(9):704–712, 2020.

[10] Ivan A Moreno-Hernandez, Michelle F Crook, Vida Jamali, and A Paul Alivisatos. Recent advances in the study of colloidal nanocrystals enabled by in situ liquid-phase transmission electron microscopy. *MRS Bulletin*, 47(3):305–313, 2022.

[11] Zihao Ou, Ziwei Wang, Binbin Luo, Erik Luijten, and Qian Chen. Kinetic pathways of crystallization at the nanoscale. *Nature materials*, 19(4):450–455, 2020.

[12] Isabel Panicker, Zain Shabeeb, Cory Hargus, and Vida Jamali. Modulating nanoparticle-surface interactions at liquid-solid interfaces via ion screening in liquid-phase tem. *ChemRxiv*, 2025. doi: 10.26434/chemrxiv-2025-60c5k.

[13] Federico Perazzi, Anna Khoreva, Rodrigo Benenson, Bernt Schiele, and Alexander Sorkine-Hornung. Learning video object segmentation from static images. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2663–2672, 2017.

[14] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alex Sorkine-Hornung, and Luc Van Gool. The 2017 davis challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*, 2017.

[15] Luis Rangel DaCosta, Katherine Sytwu, CK Groschner, and MC Scott. A robust synthetic data generation framework for machine learning in high-resolution transmission electron microscopy (HRTEM). *npj Computational Materials*, 10(1):165, 2024.

[16] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. SAM 2: Segment anything in images and videos. *The Thirteenth International Conference on Learning Representations*, 2024.

[17] Kevin M Roccapriore, Ondrej Dyck, Mark P Oxley, Maxim Ziatdinov, and Sergei V Kalinin. Automated experiment in 4d-stem: exploring emergent physics and structural behaviors. *ACS nano*, 16(5):7605–7614, 2022.

[18] Frances M Ross. Opportunities and challenges in liquid cell electron microscopy. *Science*, 350 (6267):aaa9886, 2015.

[19] Martin Schorb, Isabella Haberbosch, Wim JH Hagen, Yannick Schwab, and David N Mastronarde. Software tools for automated transmission electron microscopy. *Nature methods*, 16 (6):471–477, 2019.

[20] Zain Shabeeb, Naisargi Goyal, Pagnaa Attah Nantogmah, and Vida Jamali. Learning the diffusion of nanoparticles in liquid phase tem via physics-informed generative AI. *Nature Communications*, 16(1):6298, 2025.

[21] Katherine Sytwu, Catherine Groschner, and Mary C Scott. Understanding the influence of receptive field and network complexity in neural network-guided tem image analysis. *Microscopy and Microanalysis*, 28(6):1896–1904, 2022.

[22] Xingzhi Wang, Jie Li, Hyun Dong Ha, Jakob C Dahl, Justin C Ondry, Ivan Moreno-Hernandez, Teresa Head-Gordon, and A Paul Alivisatos. Autodetect-mnp: an unsupervised machine learning algorithm for automated analysis of transmission electron microscope images of metal nanoparticles. *JACS Au*, 1(3):316–327, 2021.

[23] Taylor J Woehl, Trevor Moser, James E Evans, and Frances M Ross. Electron-beam-driven chemical processes during liquid phase transmission electron microscopy. *MRS Bulletin*, 45(9): 746–753, 2020.

[24] Lehan Yao, Zihao Ou, Binbin Luo, Cong Xu, and Qian Chen. Machine Learning to Reveal Nanoparticle Dynamics from Liquid-Phase TEM Videos. *ACS Central Science*, 6(8):1421–1430, 2020. ISSN 2374-7943.

[25] Ze-Wei Ye, Hung-Wei Hsueh, and Shu-Han Hsu. Diversity-based two-phase pruning strategy for maximizing image segmentation generalization with applications in transmission electron microscopy. In *AI for Accelerated Materials Design - NeurIPS 2024*, 2024.

Appendix

A   Dataset and Simulation Details

Our simulated training set consists of 932 videos each with 50 frames of size of 1024×1024 pixels, where each of the videos has water thickness uniformly randomly sampled from a distribution of $\{5, 10, 25, 50, 75, 100, 125, 150, 160\}$ nanometers (nm), with visual differences shown in Fig. 5. Each video contains a number of particles selected uniformly from a range of 1 to 8, with possible various interactions between particles, such as collisions or overlap, ensuring robustness of training.

For our testing set, we used

1. Video (1024×1024) consisting of 300 frames and 3 particles, including occasional frame-to-frame overlaps, used for general performance evaluation (Fig. 2)

2. Video (512×512) consisting of 275 frames and a single particle, providing a lower-resolution test case (Fig. 4)

3. Video (1024×1024) consisting of 600 frames and 5 particles, with minimal interactions, used to evaluate tracking accuracy and mask stability over long trajectories (Fig. 7).
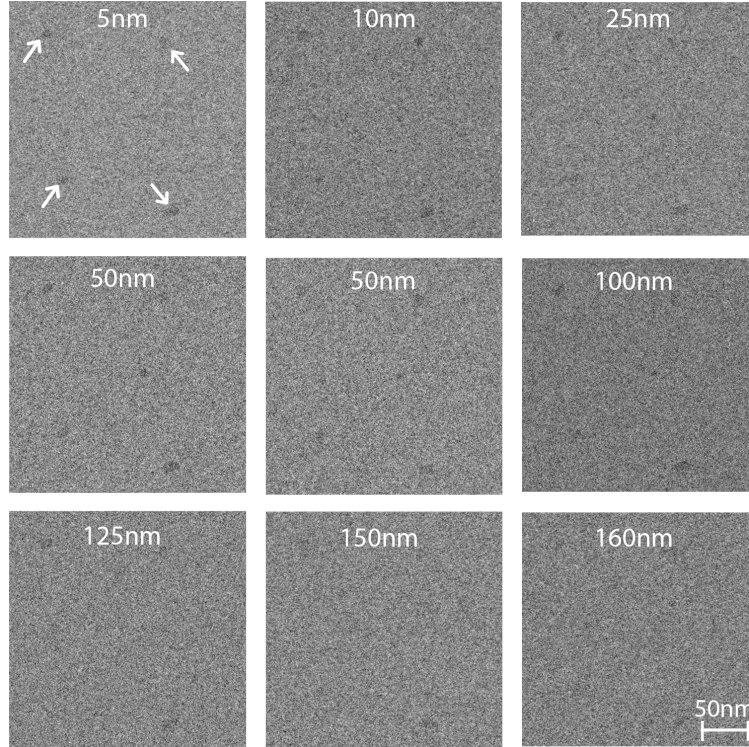

Figure 3: Example simulated LPTEM frames at different liquid thicknesses, illustrating the decrease in image SNR as thickness (and electron scattering) increase.

A.1   Accuracy Metrics

The accuracy of the segmentation on the simulated single-particle datasets was evaluated using the $\mathcal{J}$ & $\mathcal{F}$ metric [16] on a scale from 0% to 100%, where a score of 100% indicates that the prediction completely matches the ground truth. The Jaccard Index $\mathcal{J}$ assesses region similarity by calculating the intersection-over-union between the segmented mask (M) and the ground-truth (G): $\mathcal{J} = \frac{|M \cap G|}{|M \cup G|}$. The F-measure $\mathcal{F}$ assesses the accuracy of the mask contour. $\mathcal{F}$ is calculated from the contour-based precision (P_c) and recall (R_c) between the boundary pixels of the segmented mask
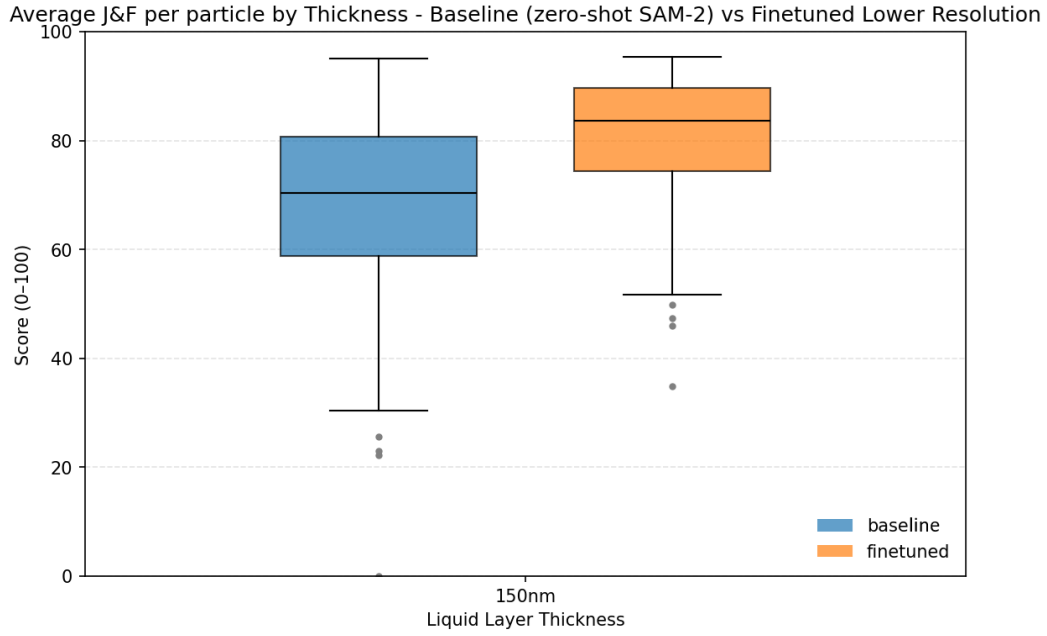
Figure 4: Performance accuracy of SAM-EM in segmentation on a simulated dataset with resolution of 512×512 resolution. Comparison of fine-tuned SAM-EM model and baseline (zero-shot SAM 2) on a simulated test dataset with 275 frames with one moving nanoparticle.
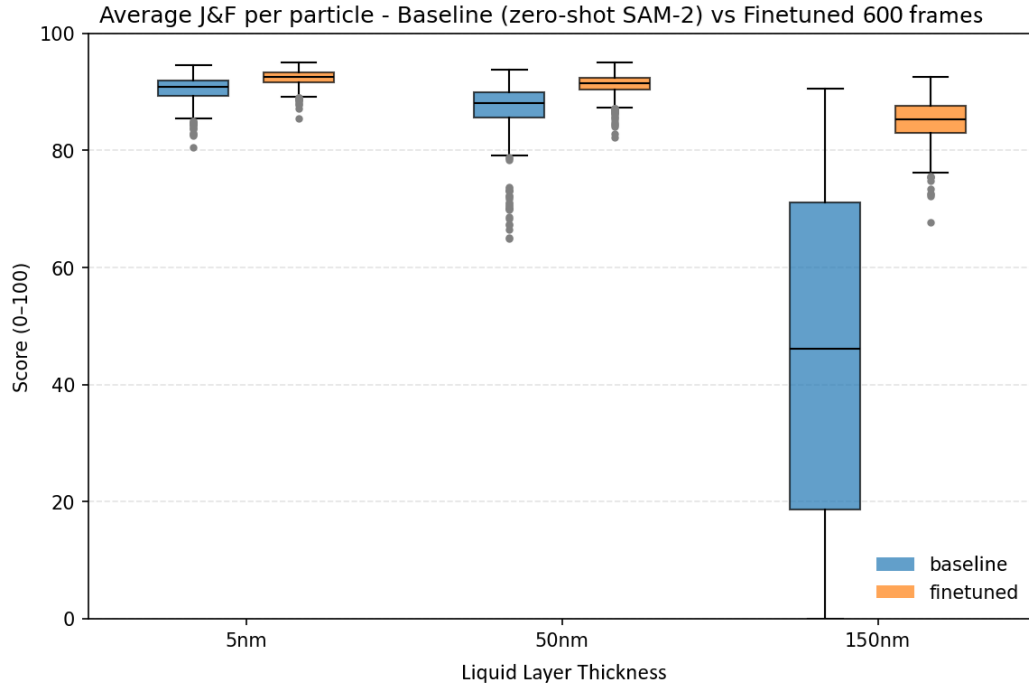

Figure 5: Performance accuracy of SAM-EM in segmentation on simulated datasets with resolution of 1024×1024. Comparison of fine-tuned SAM-EM model and baseline (zero-shot SAM 2) on a simulated test dataset with 600 frames with four moving nanoparticles and with minimal particle interaction.

and the ground-truth: $\mathcal{F} = \frac{2*P_c*R_c}{P_c+R_c}$. See [13] for further details on the $\mathcal{J}$ and $\mathcal{F}$ metrics. The $\mathcal{J}$ & $\mathcal{F}$ score of a frame is the average of the $\mathcal{J}$ and $\mathcal{F}$ of the frame [14]. The average $\mathcal{J}$ & $\mathcal{F}$ across the video is computed using the per-frame scores.

B   Fine-tuning Details

We fine-tuned the SAM 2 Hiera-Large checkpoint on synthetic LPTEM video datasets using a single NVIDIA GH200 GPU (96 GB VRAM). Training used a batch size of 1 (constrained by memory), a vision learning rate of 1×10^{-7}, a frame count of 8 (i.e., the number of consecutive frames the model was exposed to per step), and 15 epochs. These hyperparameters were selected empirically to balance stability and convergence for noisy video data. The full YAML configuration, along with training scripts, is available in our public GitHub repository.
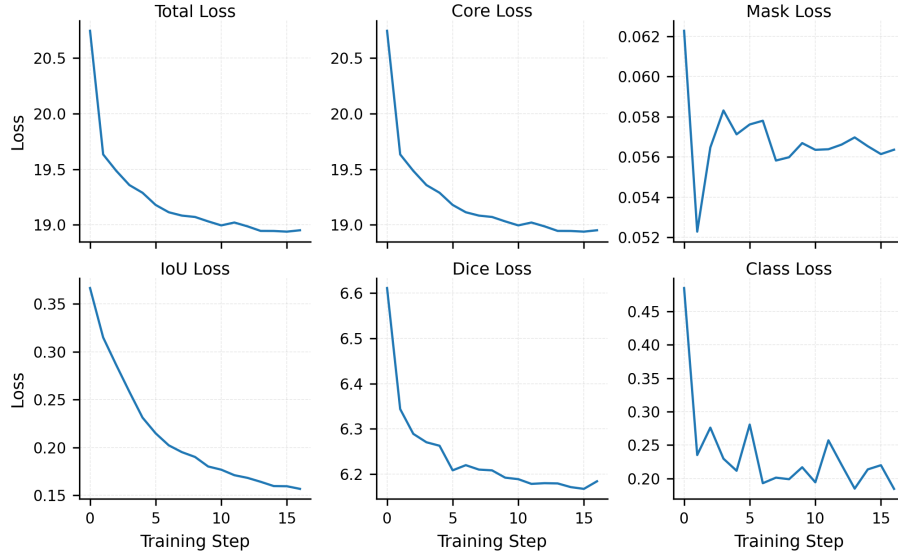
C   Fine-tuning Loss Curves

Figure 6: **Generated training loss curves during model fine-tuning**. The higher initial decline and gradual slowing of the total loss show a healthy loss curve, suggesting an appropriate learning rate

D   Statistical Analysis of Particle Trajectories

D.1   Single particle trajectory analysis

The centroid (x, y) and orientation (in-plane angle, θ) of each particle in each frame are calculated from the segmented mask array in the Particle Tracking module. The data is saved as a .csv file.

To characterize the diffusive behavior of the nanoparticles in LPTEM experiments, we analyzed the trajectories using three key statistical metrics.

D.1.1   Time-averaged Mean Squared Displacement

The time-averaged $\text{MSD}(\tau)$ quantifies the average displacement of a particle over time delay, τ. It is given by:

$$\text{MSD}(\tau) = \langle (\mathbf{r}(t+\tau) - \mathbf{r}(t))^2 \rangle,$$

where $\mathbf{r}(t)$ and $\mathbf{r}(t+\tau)$ represent the particle's position at time t and $t + \tau$, respectively. $\langle \cdot \rangle$ denotes the mean value over frames of a single trajectory.

D.1.2 Velocity Autocorrelation

The velocity autocorrelation function, $C_{\mathbf{v}}(\tau)$, measures the temporal memory of a particle's motion and is defined as:

$$C_{\mathbf{v}}(\tau) = \frac{\langle \mathbf{v}(t) \cdot \mathbf{v}(t+\tau) \rangle}{\langle \mathbf{v}(t)^2 \rangle},$$

where $\mathbf{v}(t)$ and $\mathbf{v}(t+\tau)$ represent the particle's velocities at time t and $t+\tau$, respectively. This function indicates whether the motion is correlated, uncorrelated, or anti-correlated.

D.1.3 Distribution of Displacements

The distribution of displacements provides statistical information about the nature of the particle's motion. This is characterized by the probability density function (PDF) of particle displacements $\Delta \mathbf{r}$ over a fixed time interval. For Brownian motion, the distribution is Gaussian, while non-Gaussian processes may exhibit heavy tails or other deviations, reflecting the underlying physical mechanisms influencing the motion.
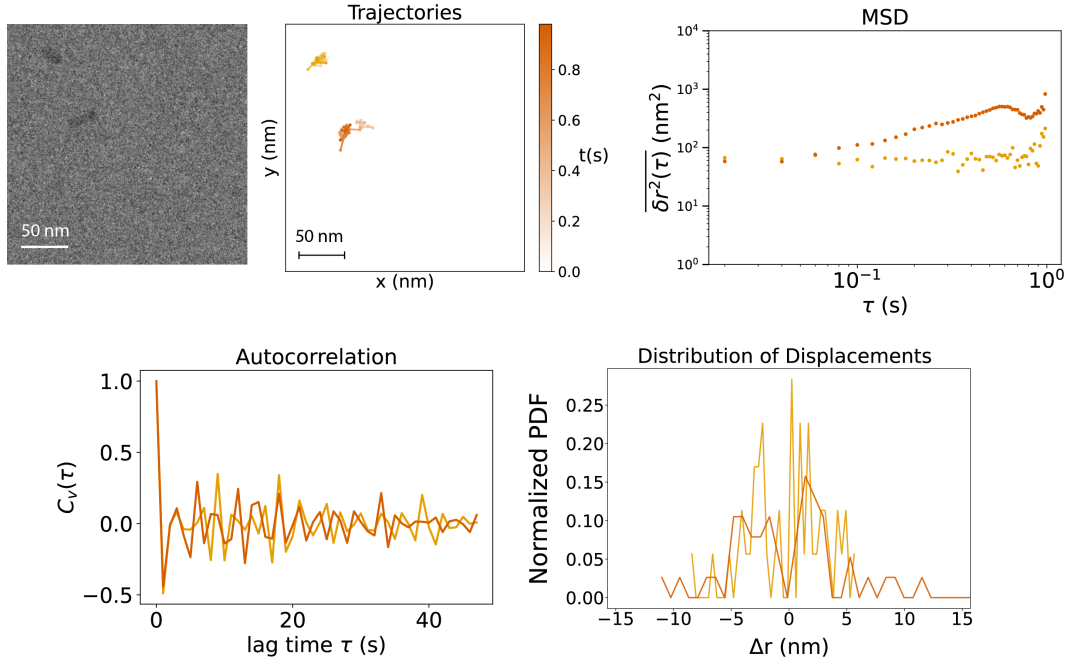

Figure 7: Statistical analysis and trajectory plots of two particles in high-noise LPTEM simulated data (50 frames, 5nm liquid layer), demonstrating that SAM-EM maintains accurate tracking and quantitative characterization (MSD, distribution of displacement, and velocity autocorrelation) under low-SNR conditions.

E Real Experimental Data Evaluation

E.1 Qualitative Analysis

To showcase the power of SAM-EM on real LPTEM data, we applied it on an experimental LPTEM video (200 frames) acquired at an electron dose rate of 35 e$^-$/Å^2·s and at a frame rate of 80 fps, equivalent to 0.0125 seconds exposure time. For this low-contrast dataset, SAM-EM generates coherent ellipsoidal masks that align closely with manually annotated particle masks. The largest discrepancies appear along particle boundaries, where hand-annotated masks tend to be slightly thicker or rotated, an expected deviation given both the human annotation error and the weak signal at nanoparticle edges. Visual inspection confirms that centroid positions remain well-aligned across all four particles, even when mask sizes differ.
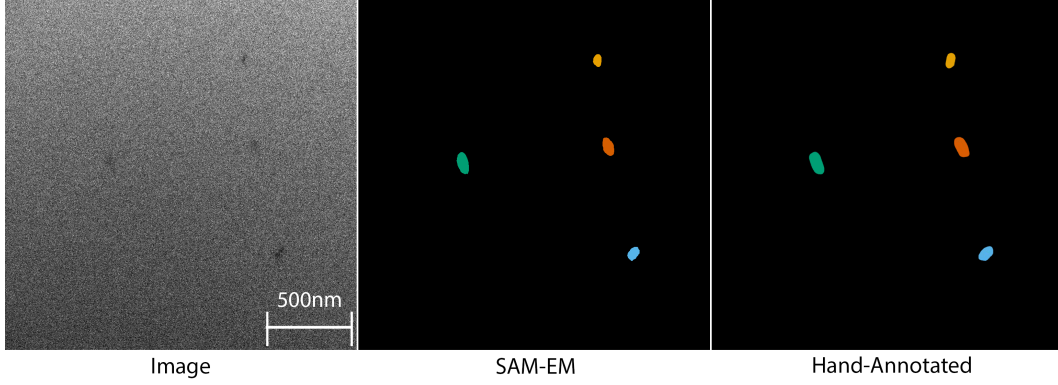
11

Figure 8: Real LPTEM frame with hand-annotated labels and SAM-EM segmentation masks. Left: raw image. Middle: SAM-EM segmentation (one box prompt per particle). Right: hand-annotated masks used as *pseudo*-ground truth for visualization. SAM-EM identifies the same four nanoparticles with compact, identity-consistent masks. Apparent boundary differences largely reflect label imprecision under low SNR rather than missed detections.

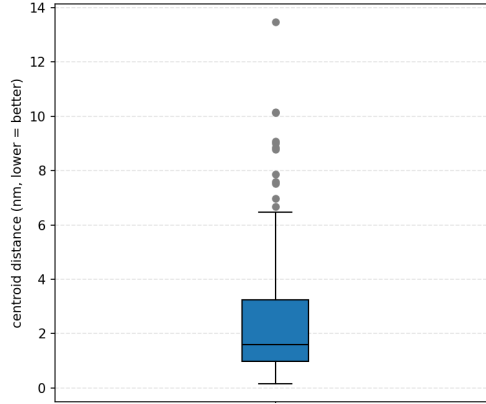## E.2 Centroid Agreement Analysis

Figure 9: Box plot of centroid distances (nm) between SAM-EM predictions and manually annotated particle centers across 200 frames from a real LPTEM dataset. The median offset of 1.8 nm is within the expected uncertainty of manual annotation under low-SNR imaging conditions.

Because hand-annotated data in LPTEM is inherently noisy and inconsistent at the pixel level, we evaluated SAM-EM using centroid agreement rather than mask accuracy. Under low SNR conditions, the nanoparticle boundaries are often subjective and can be interpreted differently. Different annotators may draw slightly thicker or rotated masks even when identifying the same particle, leading to misleading results if mask accuracy is used directly for evaluation. As shown in Fig. 9, SAM-EM achieves a median centroid deviation of ≈1.8 nm relative to hand-annotated masks. The narrow interquartile range indicates consistent alignment between model predictions and manual annotations, while the presence of outliers reflects regions where boundaries are highly occluded or ambiguous. This analysis demonstrates that SAM-EM reliably localizes particle centers with sub-particle precision, making it suitable for downstream single-particle trajectory analysis and quantitative studies of nanoscale motion. We note that, unlike the simulated data in previous figures, experimental datasets lack a true ground truth; thus, Fig. 9 reports only the distance between centroids in SAM-EM and manually segmented masks.

F   Resources

All of our code is available on GitHub: github.com/JamaliLab/SAM-EM.

Code for simulated data generation: github.com/JamaliLab/LPTEMsimulator.

All datasets and model checkpoints are available on Hugging Face: https://huggingface.co/sam-em-paper