Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Turning Large Electron Microscopy Datasets into Quantitative Measurements with Model-Assisted Analysis

Alexis N. Williams

williamsan@ornl.gov

Oak Ridge National Laboratory https://orcid.org/0000-0002-5283-5822

Chanda R. Harris

Oak Ridge National Laboratory

Juhyun Oh

Oak Ridge National Laboratory https://orcid.org/0000-0002-4492-6978

Michael J. Zachman

Oak Ridge National Laboratory https://orcid.org/0000-0003-1910-1357

Albina Y. Borisevich

Oak Ridge National Laboratory https://orcid.org/0000-0002-3953-8460

Eric R. Hoglund

Oak Ridge National Laboratory https://orcid.org/0000-0002-3409-4960

Asraful Haque

Oak Ridge National Laboratory https://orcid.org/0009-0004-7640-6784

Sumner B. Harris

Oak Ridge National Laboratory https://orcid.org/0000-0003-4905-0482

Miguel Toro-González

Oak Ridge National Laboratory https://orcid.org/0000-0003-3320-423X

Jack C. Lasseter

Oak Ridge National Laboratory https://orcid.org/0000-0001-8231-1125

Leah H. Hochanadel

Oak Ridge National Laboratory https://orcid.org/0009-0006-9282-0832

Delyana P. Vasileva

Oak Ridge National Laboratory https://orcid.org/0000-0002-7919-7883

Jennifer L. Morrell-Falvey

Oak Ridge National Laboratory https://orcid.org/0000-0002-9362-7528

Joshua K. Michener

Oak Ridge National Laboratory https://orcid.org/0000-0003-2302-8180

Sandra M. Davern

Oak Ridge National Laboratory https://orcid.org/0000-0003-2755-8912
Brian C. Sanders
Oak Ridge National Laboratory https://orcid.org/0000-0001-9754-1687
Scott T. Retterer
Oak Ridge National Laboratory https://orcid.org/0000-0001-8534-1979

Research Article

Keywords: cryogenic electron microscopy, low-dose imaging, image segmentation and machine learning, data provenance, simulation, model training

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11512593/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

DOE PUBLIC ACCESS NOTICE

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (`https://www.energy.gov/doe-public-access-plan`).

Turning Large Electron Microscopy Datasets into Quantitative Measurements with Model-Assisted Analysis

Alexis N. Williams[1,*], Chanda R. Harris[1], Juhyun Oh[1], Michael J. Zachman[1], Albina Y. Borisevich[1], Eric R. Hoglund[1], Asraful Haque[1], Sumner B. Harris[1], Miguel Toro-González[4], Jack C. Lasseter[1], Leah H. Hochanadel[3], Delyana P. Vasileva[3], Jennifer L. Morrell-Falvey[3], Joshua K. Michener[3], Sandra M. Davern[2], Brian C. Sanders[3], and Scott T. Retterer[1]

[1]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

[2]Radioisotope Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

[3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

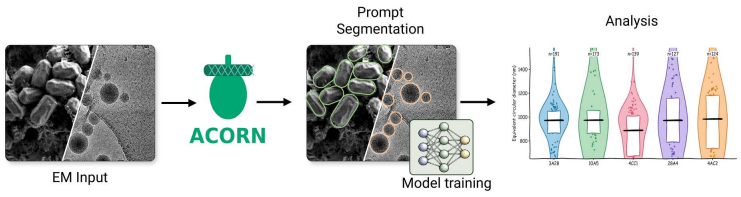
[4]NorthStar Medical Technologies Beloit, Wisconsin 53511, USA

[*]Corresponding author: williamsan@ornl.gov

Abstract

Large electron microscopy datasets require extensive annotation and review before they can be converted into reliable measurements. ACORN is an open-source, locally executable application that offers image preparation, reviewed annotation, task-specific model training, batch inference, calibrated measurement, and statistical analysis in a single format. Users can create and correct annotations manually or with promptable segmentation and use the reviewed annotations to train models directly within ACORN. Its Cryo-EM Lab Utility (CLU) assistant also allows users to request any analysis or training operations in plain language. To show its use cases, we evaluated ACORN using low-dose cryogenic TEM images of polymeric nanoparticles and SEM images of bacterial spores. YOLO models trained in ACORN produced validation F1 scores of 0.906 ± 0.013 and 0.624 ± 0.016, respectively. Controlled simulations built in ACORN provided 25,983 exact object annotations for further model development and a controlled evaluation of software performance. On held-out simulation images, equivalent-circular-diameter mean absolute error was 3.60 nm (10.1%) for TEM particles and 31.48 nm (3.7%) for SEM spores among matched detections. The software logs every interaction, training step, and annotation in a sample history for provenance tracking. ACORN provides an environment where annotations, image calibration, model checkpoints, object measurements, and statistical records are easily accessible and traceable. This software provides a traceable path from electron microscopy images to quantitative results in multiple user-friendly formats.

Keywords: cryogenic electron microscopy; low-dose imaging; quantitative imaging; image segmentation and machine learning; data provenance; simulation; model training



1 Introduction

In the life sciences, automated electron microscopy (EM) generates images faster than researchers can convert them into quantitative results. As EM automation reaches into materials science, a similar analysis bottleneck is occurring. A dataset may be acquired in hours with little user input, whereas identifying objects of interest, tracing their boundaries, converting pixel-based measurements into physical units, and comparing measurements across experimental conditions can require days of analysis. Manual annotation remains suitable for small datasets but becomes a problem when studies contain hundreds of images or thousands of objects. Automated segmentation can reduce this burden as long as these predictions can incorporate image calibration, annotation history, morphological measurements, and statistical analysis.

General computer-vision platforms use an iterative annotation and training process. Users label an initial subset of images, train a model, apply that model to unlabeled images, correct its predictions, and retrain it using the expanded dataset. Roboflow, CVAT, and Label Studio support different stages of this process through browser-based or self-hosted interfaces [1, 2, 3]. Applying this process to EM is difficult because the objects of interest vary widely among modalities and may include biological particles, cells, fibers, pores, precipitates, nanoparticles, atomic columns, or interfaces. Image contrast may arise from electron scattering, phase contrast, diffraction, composition, or detector response. Low-dose transmission electron microscopy (TEM) introduces additional processing requirements, including frame selection, motion correction, dose weighting, and contrast-transfer correction. Specialized particle-picking packages support structural-biology analyses, while cell-segmentation tools cover biological imaging more broadly [4, 5, 6, 7]. However, the outputs for these processes are generally designed for specific downstream tasks, such as coordinate files for single-particle analysis or masks in pixel units for cell counting.

For EM applications outside of established structural-biology protocols, researchers often combine general-purpose annotation, model-training, image-processing, measurement, and statistical software from varying programs. Microscopy Image Browser, μSAM, AtomAI, and DL4MicEverywhere are some examples of software for segmentation, model training, quantitative analysis, and reproducible execution [8, 9, 10, 11].Moving between these programs can separate annotations from calibration, processing history, and source images. In the case of upload-dependent platforms, these may also be unsuitable for large, proprietary, unpublished, or export-controlled datasets.

Promptable segmentation models can accelerate initial annotation of EM data. However, the masks still require review and performance may not generalize across EM modalities given the initial training in these cases commonly occur on non-EM images. For users wanting repeated analysis of a particular specimen type, these foundational-based prompt masks can feed into training a smaller task-specific model. Training such a model commonly requires scripting, dataset formatting, and hyperparameter selection that places the process beyond the experience of many microscopists. EM researchers need accessible, iterative model training in software that retains calibration, provenance, and quantitative measurements.

An additional limitation with EM is that labeled datasets are scarce. For these cases, simulation micrographs offer a possible starting point because object locations, boundaries, and properties are known. These simulations used for model training or evaluation must capture enough variation in specimen structure, acquisition conditions, contrast, and noise to support transfer to experimental images [12, 13, 14, 15, 16, 17, 18, 19, 20]. ACORN (Annotate, Curate, Observe, Review, and Navigate) is an open-source application for local EM analysis. It reads microscopy data in its native format, preserves calibration from image headers, supports multiple annotation methods, trains task-specific models, and calculates measurements and statistics from large datasets. CLU (Cryo-EM Lab Utility) converts natural-language requests into registered ACORN operations and records

those actions in the session history. End users can develop their own task-specific capabilities and integrate those into the ACORN base package via plug-ins. To showcase its abilities, we evaluated ACORN using low-dose cryo-EM micrographs of nanoparticles and SEM micrographs of bacterial spores. Figure 1 presents the software overview and use cases.

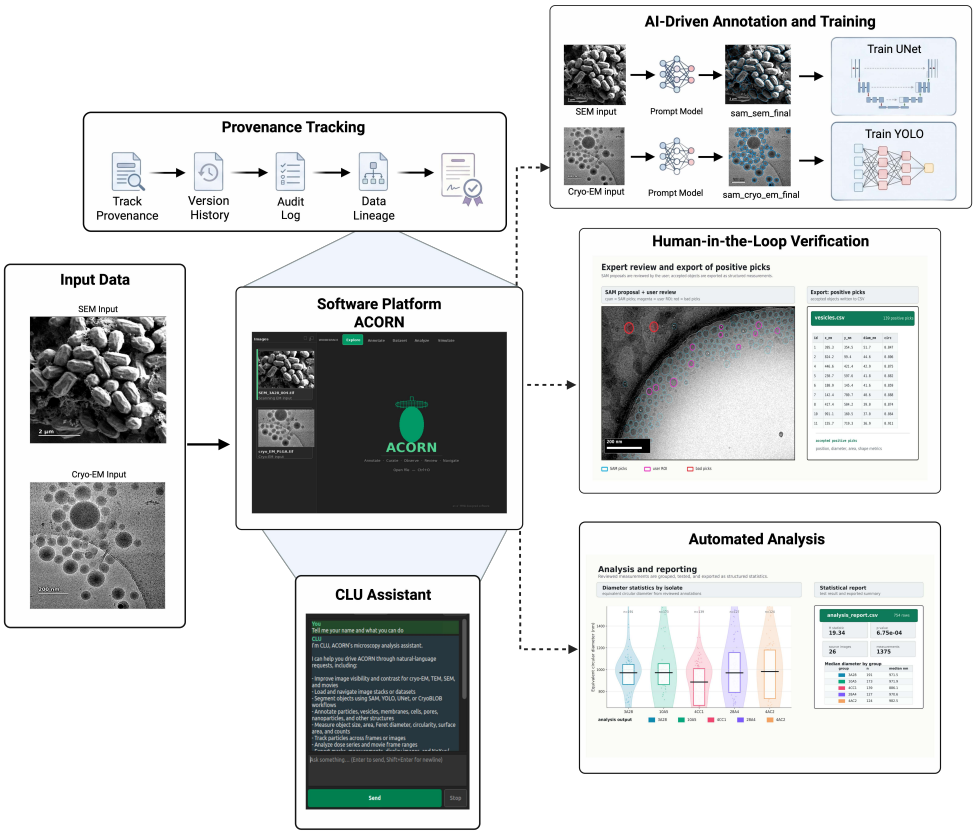


Figure 1: **ACORN connects EM images to quantitative analysis.** ACORN links microscopy images and metadata with reviewed annotations, task-specific model training, calibrated measurements, statistical analysis, and provenance records. CLU provides natural-language access to the same registered software operations.

2 Results

2.1 Demonstrating the capabilities of ACORN - image annotation, quantitative analysis, and provenance tracking

ACORN is a Python application of 51,103 lines distributed as eight installable packages that share a dedicated core install. End users can choose from the desktop application for interactive analysis or a headless command-line interface for image conversion, batch prediction, dataset export, and dataset finalization. ACORN includes the Cryo-EM Lab Utility (CLU), an in-application natural-language assistant that can access all 57 registered tools for data loading, contrast adjustment, segmentation, measurement, plotting, model training, export, and simulation. Image processing and

model inference run on various compute setups in an OS agnostic format. It can use either local CPU or GPU resources or run on cluster-based systems with GPU parallelization when needed. CLU can use locally hosted language models for isolated network deployment or a cloud-hosted model through an API key. Supplementary Table S9 compares ACORN with general-purpose annotation platforms and microscopy software.

A benefit of using ACORN for processing and training is spatial calibration remains associated with the imported datasets, unlike other instances where this pixel size is lost. The software records the pixel size and whether it came from instrument metadata, in-application calibration, or manual entry. ACORN marks images lacking spatial calibration as 'uncalibrated' or 'pixel size unknown' and automatically adjusts the calibrated pixel size after operations that may modify apparent size (e.g., binning). Across 613 post-correction automated tests spanning image import, annotation, metadata restoration, binning, measurement, and export, ACORN retained pixel size with zero relative error (Supplementary Table S11). Furthermore, ACORN records processing choices that can affect object boundaries and alerts the user to the corresponding measurement limitations. This means when a user measures a feature of interest with any annotation method in the software the output sizes are appropriately calibrated independent of any pre-processing steps. Measurement and statistical records retain experimental units, grouping and exclusion variables, statistical tests, significance thresholds, correction methods, effect sizes, and exact p values. ACORN associates reported measurements and plotted values to their annotations, source images, spatial calibrations, and analysis parameters for easy replication. To show this we audited and reported the stored origins of all 1,375 final SEM annotation records across 26 source micrographs, comprising 1,085 imported annotations (78.9%), 51 prompt-generated annotations (3.7%), and 239 trained-model annotations (17.4%) (Supplementary Figure S9). We then traced two representative analysis chains through stored identifiers and hashes to document the connections among source images, annotations, model checkpoints, measurements, and downstream results (Supplementary Table S12).

2.2 Motion correction and dose binning allows for measurement of dynamics

Structural biology cryo-EM routinely uses motion correction of dose-fractionated image stacks [21, 22, 23]. ACORN implements two-pass rigid phase cross-correlation for these stacks [24]. It preserves the frames on import so users can inspect the acquisition, select a frame range, estimate and remove coherent image drift, and construct dose bins from the aligned frames. ACORN displays all dose-bin averages and first-bin differences with common intensity scales and reports the recovered drift trajectory, frame-to-frame displacement, maximum drift, and total path length in pixels and, when calibration is available, physical units. Motion correction and dose binning allow researchers to remove global sample displacement that blurs the same structure across frames, while binning trades temporal or dose resolution for signal-to-noise ratio.

We evaluated these functions using a 100-frame simulation of a nanoparticle-loaded polymer thin film with programmed global drift, local membrane deformation, nanoparticle motion relative to the membrane, and dose-dependent contrast loss (Figure 2).

The simulation contained 100 frames of 512×512 pixels at 0.5 nm/pixel. Each frame represented $10\ e^{-}\ \text{Å}^{-2}$ giving a total accumulated dose of $1000\ e^{-}\ \text{Å}^{-2}$. Global image translation was programmed independently from local membrane deformation and nanoparticle motion relative to the membrane (Figure 2A). ACORN recovered the global drift trajectory with a RMSE of 0.61 pixels (Figure 2B). Alignment reduced the coherent motion blur in the uncorrected frame average (Figure 2C). The aligned frames were then grouped by cumulative dose. The 100 $\text{e}^{-}\ \text{Å}^{-2}$ bins retained the membrane and nanoparticle structure across the simulated dose series independent of dose (Figure 2D).

Dose-bin width was used to balance image contrast and dose resolution (Figure 2E). At a

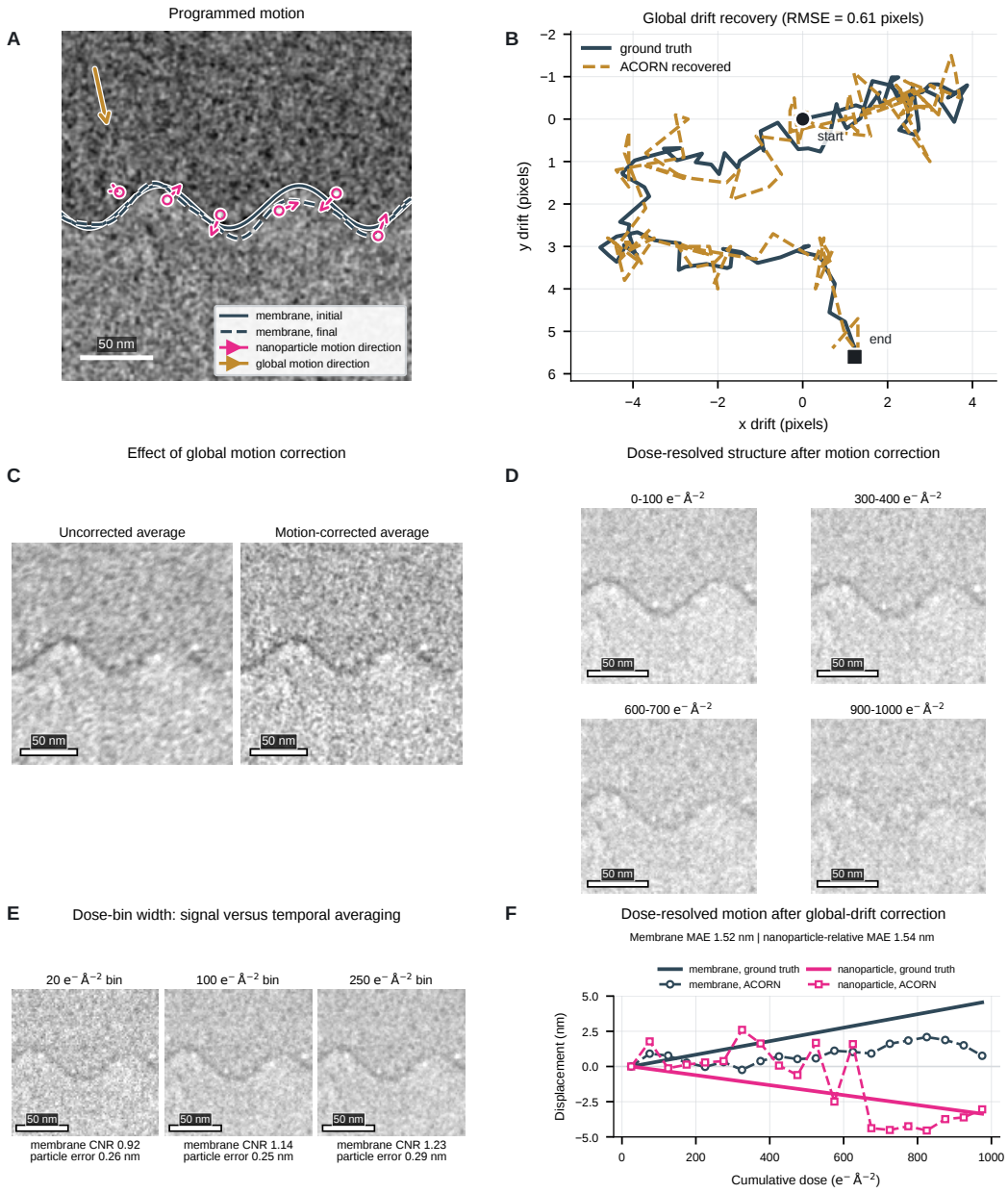


Figure 2: **Motion correction and dose binning recover global drift and dose-resolved local motion.** (A) Programmed motion in a simulated nanoparticle-loaded polymer thin film. The solid and dashed contours show the initial and final membrane positions, respectively. Gold and magenta arrows indicate global image motion and nanoparticle motion relative to the membrane. (B) Programmed and ACORN-recovered global drift trajectories RMSE = 0.61 pixels. (C) Average of the uncorrected frames compared with the average after rigid motion correction. (D) Motion-corrected structure in four 100 $e^- Å^{-2}$ dose intervals spanning a cumulative dose of 0–1000 $e^- Å^{-2}$. (E) Comparison of dose-bin widths centered near a cumulative dose of 500 $e^- Å^{-2}$. Increasing the bin width from 20 to 250 $e^- Å^{-2}$ increased membrane contrast-to-noise ratio. (F) Programmed and recovered membrane displacement and nanoparticle displacement relative to the local membrane after global-drift correction.

cumulative dose near 500 e^{-} Å^{-2}, increasing the bin width from 20 to 100 e^{-} Å^{-2} increased the membrane contrast-to-noise ratio from 0.92 to 1.14, while nanoparticle-localization error changed from 0.26 to 0.25 nm. Increasing the bin width to 250 e^{-} Å^{-2} raised the membrane contrast-to-noise ratio further to 1.23 but increased localization error to 0.29 nm and averaged motion over a longer dose interval. Among the three tested settings, the 100 e^{-} Å^{-2} bin gave the lowest nanoparticle-localization error while retaining greater membrane contrast than the 20 e^{-} Å^{-2} bin.

After global-drift correction ACORN separately tracked displacement of the membrane and the nanoparticles relative to the local membrane (Figure 2F). The recovered membrane and nanoparticle-relative trajectories had mean absolute errors of 1.52 and 1.54 nm, respectively. A benchmark used 48 simulated movies, with six movies per dose condition and independently generated specimen configurations, drift trajectories, and noise realizations. Supplementary Figure S11 and Supplementary Table S14 report the alignment errors and the low-dose fail points.

2.3 ACORN ties segmentation to quantitative analysis

We evaluated the image import-to-result process using low-dose cryo-EM images of polymeric [poly(lactic-co-glycolic acid)-PLGA] nanoparticles in vitreous ice and SEM images of uncoated bacterial spores. For both datasets ACORN read pixel sizes from instrument metadata, associated reviewed masks with calibrated image frames, trained task-specific models, and returned object measurements in physical units. We then used CLU to automatically assign source micrographs to training, model-selection, and validation partitions and pick model training parameters. Micrographs used during prompt development were assigned to the training partition, and each model was trained with three seeds. Supplementary Table S10 reports dataset composition, partition sizes, hyperparameters, timing, and performance metrics. We showed this capability with both experimental and simulated micrographs. Supplementary Figure S8 presents predictions on representative validation micrographs.

2.4 Cryo-EM of nanoparticles

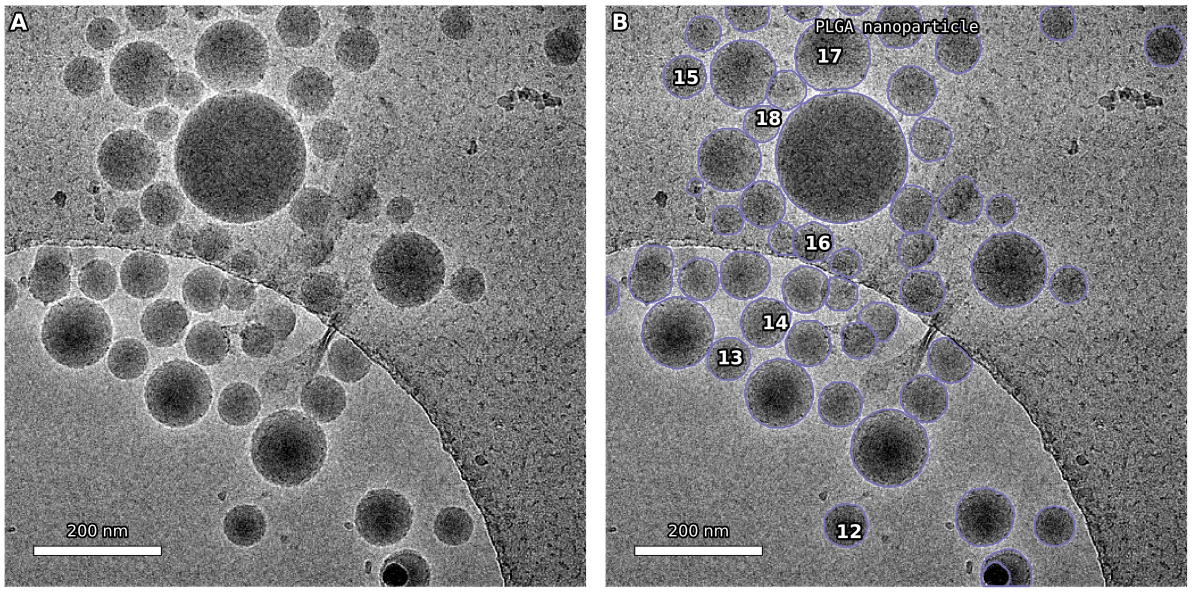
The cryo-EM dataset contained 30 micrographs and 881 reference annotations, as shown in Figure 3. We assigned 19 micrographs to training, three to model selection, and eight to validation. Precision, recall, and F1 on the validation dataset were 0.893 ± 0.027, 0.920 ± 0.009, and 0.906 ± 0.013, respectively. Mask mAP_{50} was 0.957, and mask $\mathrm{mAP}_{50\text{–}95}$ was 0.775.

We generated the initial reference masks using SAM-3. An expert manually reviewed and corrected these masks using ACORN. For an orthogonal assessment of particle-recovery performance, we evaluated the model against an independently generated CryoBlob selection [25] containing 235 particles of the same type. The model recovered 0.909 ± 0.010 of the CryoBlob-selected particles.

2.5 SEM of bacterial spores

The SEM dataset contained 26 source micrographs and 1,085 selected objects as shown in Figure 4. We assigned 14 micrographs to training, four to model selection, and eight to validation. Precision, recall, and F1 were 0.750 ± 0.001, 0.534 ± 0.023, and 0.624 ± 0.016, respectively. Mask mAP_{50} was 0.580, and mask $\mathrm{mAP}_{50\text{–}95}$ was 0.355.

In this instance, we examined whether modification of image-intensity contributed to the lower SEM performance by repeating the analysis with the annotations, source-micrograph partitions, hyperparameters, and training seeds held constant. The original per-image 0.5–99.5% percentile stretch placed 68.1% of pixels below a digital number of 32 (DN 32), with values ranging from 0



C

Pick	Object	Area (nm²)	ECD (nm)	Feret (nm)	Circularity
12	Nanoparticle 12	3,654	68.2	76.0	0.777
13	Nanoparticle 13	3,705	68.7	76.6	0.748
14	Nanoparticle 14	4,848	78.6	108.5	0.520
15	Nanoparticle 15	3,715	68.8	77.7	0.760
16	Nanoparticle 16	3,312	64.9	75.3	0.753
17	Nanoparticle 17	10,828	117.4	129.1	0.753
18	Nanoparticle 18	2,288	54.0	61.3	0.772

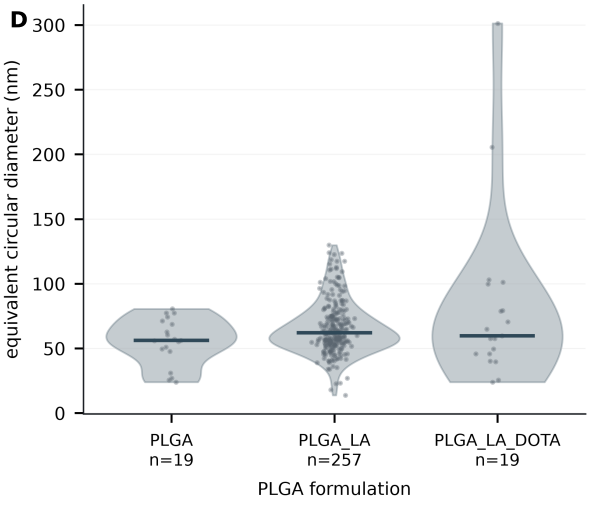
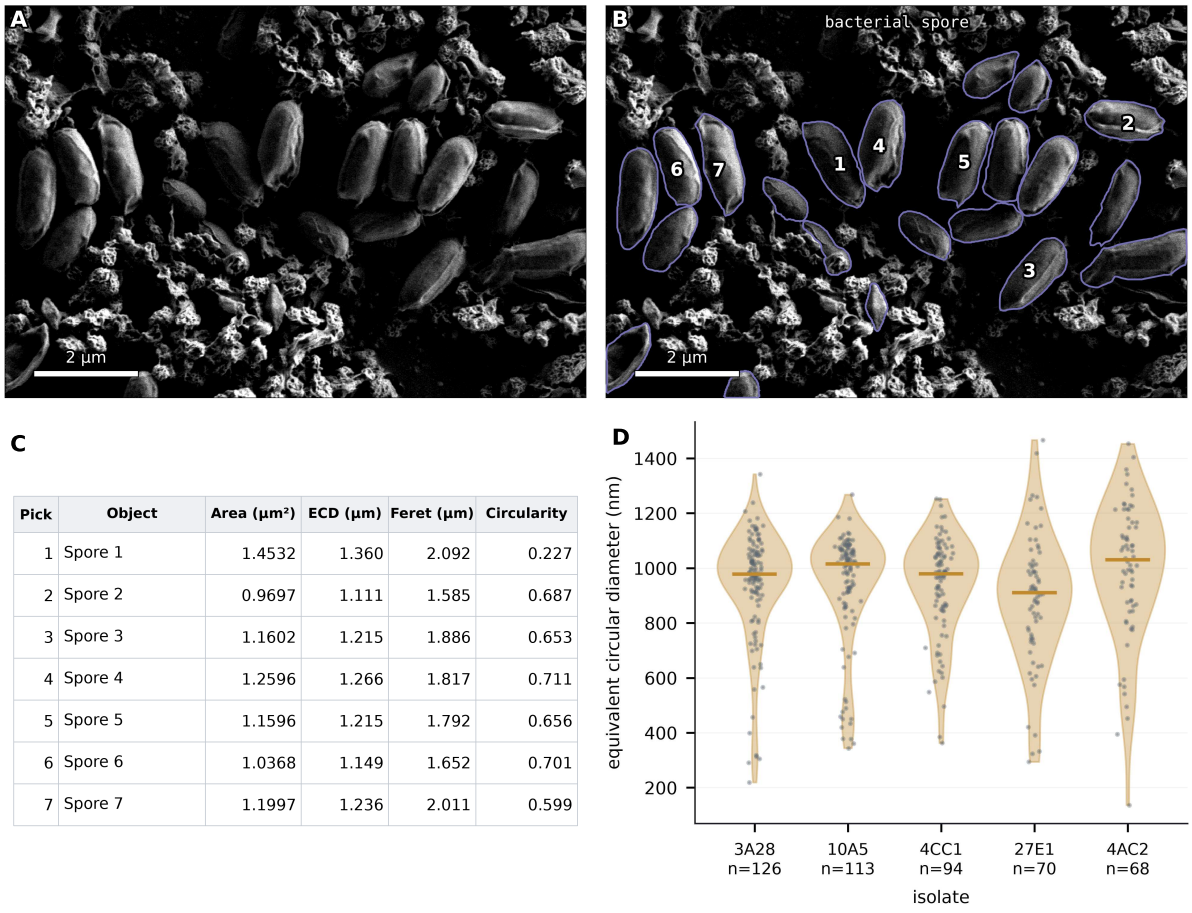


Figure 3: **Cryo-EM nanoparticle analysis in ACORN.** (A) Experimental Cryo-EM micrograph with calibration read from the MRC header. (B) Regions obtained using the prompt `PLGA nanoparticle`, corresponding to dark round objects. (C) A subset of calibrated measurements exported by the software for the detected nanoparticles. (D) Equivalent circular diameter for three PLGA formulations, analyzed using a two-sided Kruskal–Wallis test ($p = 0.259$, $n = 295$ objects, $\alpha = 0.05$). These object-level comparisons are provided to display ACORN's measurement and statistical-analysis outputs and are not intended to support specimen-level conclusions.



C

Pick	Object	Area (µm²)	ECD (µm)	Feret (µm)	Circularity
1	Spore 1	1.4532	1.360	2.092	0.227
2	Spore 2	0.9697	1.111	1.585	0.687
3	Spore 3	1.1602	1.215	1.886	0.653
4	Spore 4	1.2596	1.266	1.817	0.711
5	Spore 5	1.1596	1.215	1.792	0.656
6	Spore 6	1.0368	1.149	1.652	0.701
7	Spore 7	1.1997	1.236	2.011	0.599

Figure 4: **SEM bacterial-spore analysis in ACORN.** (A) Experimental SEM micrograph with pixel size read from the Zeiss CZ_SEM metadata. (B) Regions obtained using the prompt `bacterial spore` which corresponds to oval object. (C) Representative calibrated measurements exported for individual objects. (D) Equivalent circular diameter for the five most populous isolate codes parsed from file names. The omnibus comparison used a two-sided Kruskal–Wallis test ($p = 0.0079$, $n = 471$ objects, $\alpha = 0.05$). These object-level comparisons are provided to display ACORN's measurement and statistical-analysis outputs and are not intended to support specimen-level conclusions.

(black) to 255 (white), and produced an F1 of 0.6239 ± 0.0158. Preserving the original 8-bit SEM intensities reduced the fraction of pixels below DN 32 to 1.9% and produced an F1 of 0.6376 ± 0.0067. Applying a single global intensity window across all 26 images resulted in 1.8% of pixels below DN 32 and an F1 of 0.6176 ± 0.0314. When we confined scoring to only the morphology-filtered confident-spore subset of the reviewed reference annotations the corresponding F1 values were 0.7299, 0.7338, and 0.7119 for the per-image percentile, raw-intensity, and global-window preparations, respectively. Preserving the original intensities increased the overall F1 by 0.0137 relative to the per-image percentile stretch. This increase accounts for only a small portion of the lower SEM performance. We next tested whether the effect varied among micrographs according to their intensity distributions.

The per-image percentile stretch showed darker SEM micrographs had lower segmentation performance. Among the eight validation micrographs, the fraction of pixels below DN 32 was strongly negatively associated with F1 (Spearman $\rho = -0.881$, $p = 0.0039$). This association was absent when the original 8-bit intensities were preserved and when a single global intensity window was applied. F1 continued to vary among micrographs under both preparations, indicating that intensity preparation accounted for only part of the field-to-field variation. Supplementary Figure S10 and Supplementary Table S15 report the complete results.

ACORN assists in analyzing two diverse datasets through image import, metadata interpretation, spatial calibration, prompted annotation, YOLO model training and inference, calibrated measurement, and statistical analysis. Figures 3 and 4 show the resulting outputs for this process.

2.6 Simulation images provide reference labels for model training

Labeled EM datasets are often too small for task-specific model training. ACORN includes TEM and SEM forward models that generate simulation micrographs with exact object labels (Figure 5A,B). Across the two datasets used in this study, we generated 18,354 labeled nanoparticles and 7,629 labeled spores, for a total of 25,983 simulation annotations. The simulator recorded the position and shape of each object during image generation and automatically created the corresponding masks. ACORN saved these masks in the same format as human-generated annotations, allowing both sources to enter the same dataset-construction and model-training process.

2.7 Verifying simulated images prior to training segmentation models

Before using the simulation images for model training, we benchmarked selected outputs of the TEM and SEM forward models (Figure 5C,D). For the SEM model, we calculated backscatter yields from 20,000 electron trajectories per material and compared them with published values at 20 keV and normal incidence (Table 1).

Table 1: **Backscatter yield at 20 keV.** Model yields were calculated from traced electron trajectories and reference values were obtained from published compilations [26, 27].

Material	η model	η reference	Relative difference
Carbon	0.059	0.060	−2.0%
Silicon	0.172	0.160	+7.5%
Copper	0.326	0.300	+8.7%
Silver	0.441	0.420	+4.9%
Gold	0.548	0.490	+11.8%

The calculated backscatter yields differed from the published reference values by less than 15%

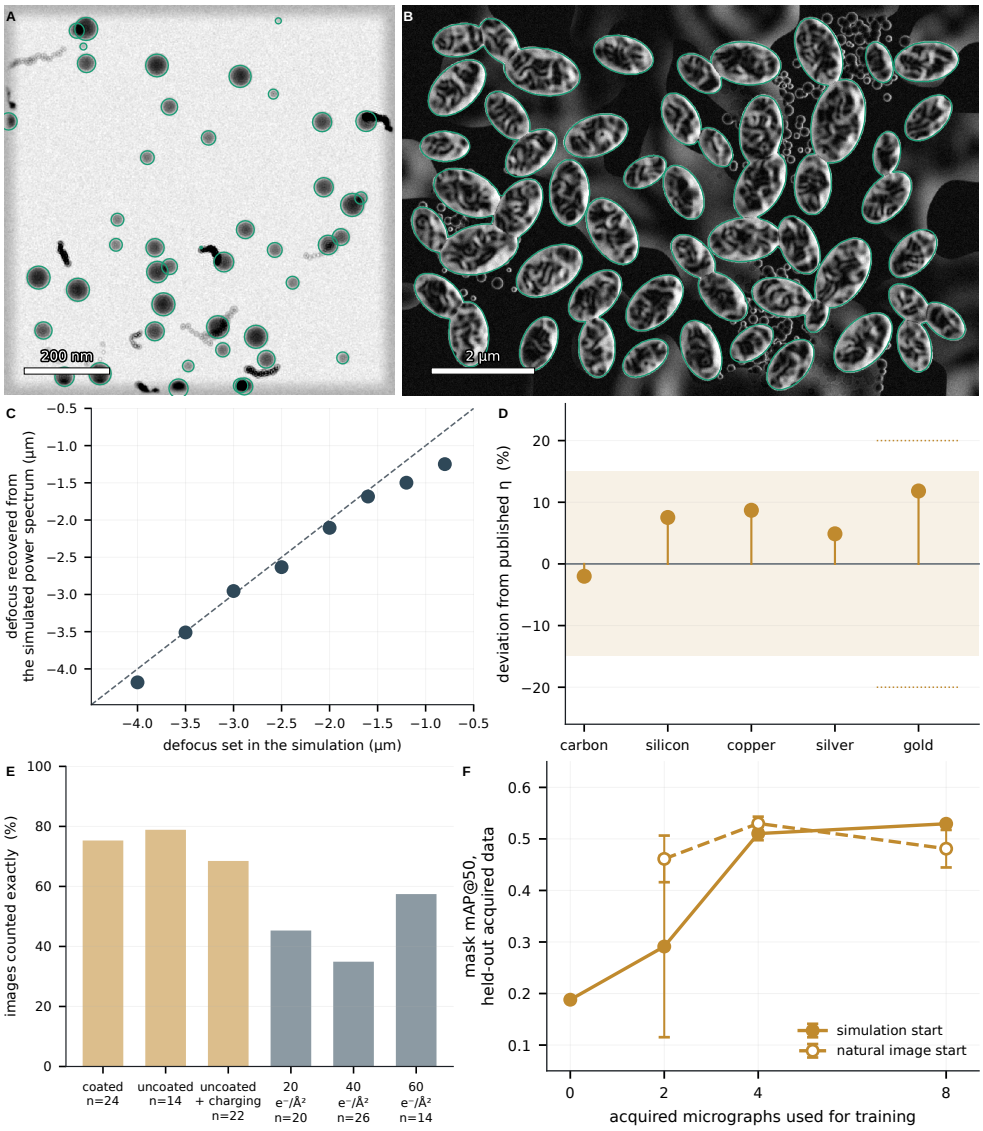


Figure 5: **Simulation, controlled benchmarks, and transfer to experimental images.** (A) Simulated cryo-EM image of PLGA nanoparticles with generated labels. (B) Simulated SEM image of bacterial spores with generated labels. (C) Transmission-model check comparing defocus recovered from the simulated power spectrum with the defocus specified in the simulation. (D) Scanning-transport check showing the deviation between calculated and published backscatter yields, η, for five materials relative to a $\pm 15\%$ band. (E) Percentage of simulated images for which the detected object count matched the known count, separated by specimen preparation for SEM and dose for TEM. (F) Mask mAP_{50} on 18 validation SEM micrographs as a function of the number of annotated experimental micrographs used for fine-tuning. Models were initialized with simulation-derived or natural-image weights. Points show mean $\pm$ 1 SD; connecting lines are visual guides.

for all five tested materials. The largest difference was +11.8% for gold, consistent with the reduced accuracy of the implemented screened Rutherford approximation for high-atomic-number materials. These results evaluate the selected outputs at 20 keV and normal incidence.

For the TEM model, we generated micrographs at specified defocus values and recovered the defocus from each image. The fit matched Thon rings in the image power spectrum to the analytical contrast-transfer phase function,

$$\chi(k) = \pi\lambda\Delta f k^2 - \frac{1}{2}\pi C_\mathrm{s}\lambda^3 k^4.$$

Fits used spatial frequencies from 0.03–0.20 /Å. Recovered and specified defocus values agreed with $R^2 = 0.986$, a mean absolute error of 0.163 μm, and a maximum error of 0.449 μm. The largest error occurred at $-0.8\,\mu$m, where relatively few Thon rings fell within the fitted frequency range. This comparison tested the internal consistency of the implemented electron wavelength, optical parameters, and contrast-transfer-function sign convention.

2.8 Controlled detection and transfer to experimental images

The final checkpoints were selected using the simulation model-selection partitions and evaluated once on 60-image test partitions. For the target class, F1 was 0.938 for TEM nanoparticles and 0.980 for SEM spores, while mask mAP_{50} was 0.935 and 0.990, respectively. Supplementary Table S6 reports the complete target-class metrics.

We varied acquisition conditions and held the simulated specimen constant (Figure 5E). We then calculated the exact-count rate as the percentage of simulated micrographs for which the detected and known object counts matched. For the SEM simulations, exact-count rates were 75% for coated spores, 79% for uncoated spores, and 68% for uncoated spores with simulated charging. For the TEM simulations, exact-count rates were 45%, 35%, and 57% at doses of 20, 40, and 60 $e^-\mathrm{Å}^{-2}$, respectively. Supplementary Figure S2 reports the complete counting-error distributions.

The exact simulated object geometries also allowed us to evaluate the mask-to-measurement path. One-to-one mask matching covered 2,097 of 2,250 TEM particles (93.2%) and 1,069 of 1,083 SEM spores (98.7%).Measurement accuracy was evaluted only for these matched true positives. Among matched detections, equivalent-circular-diameter bias and mean absolute error were +2.90 and 3.60 nm for TEM (10%) and +23.21 and 31.48 nm for SEM 3%). The remaining differences are from segmentation boundaries. Supplementary Table S16 reports the area, equivalent-circular-diameter, and Feret-diameter results.

We next evaluated whether models trained only on simulation images transferred to experimental micrographs without expert-annotated training data. Both detectors recovered nearly all objects in the held-out simulation test images (Supplementary Table S6), but their performance decreased on acquired images. The TEM detector achieved a precision of 0.970 and a recall of 0.329 against 295 verified nanoparticles. The SEM transfer results and failure modes are examined below. Supplementary Table S7 reports the complete results for each training condition. Supplementary Figure S6 presents representative predictions, while Supplementary Figures S3 and S4 present the training curves and simulated validation images, respectively.

2.9 Simulation-based model initialization

To demonstrate that ACORN can use simulated data for model initialization, we compared simulation initialization with natural-image initialization using two, four, or eight annotated experimental SEM micrographs for fine-tuning (Figure 5F). For two micrographs validation mask mAP_{50} was 0.291 $\pm$

0.176 for simulation initialization and 0.461 ± 0.045 for natural-image initialization. Performance varied substantially among the four two-micrograph subsets which shows the selected images strongly affected the result. With four micrographs the two initialization strategies produced similar values of 0.510 ± 0.013 and 0.530 ± 0.013, respectively. For all eight micrographs the simulation-initialized model produced a fixed mAP_{50} of 0.529 under deterministic training, compared with 0.481 ± 0.036 across ten natural-image-initialized executions. These results show that ACORN can use simulated images to initialize task-specific models, although simulation initialization did not provide a consistent advantage across dataset sizes. Complete replicate values and training details are reported in Supplementary Table S13.

2.10 Failure points for these methods in cryo-EM and SEM images

The cryo-EM results showed two limitations. First, the simulation-only detector transferred with high precision (0.970) but low recall (0.329) against 295 verified nanoparticles. This means most returned detections were correct, but approximately two-thirds of the reference particles were missed. Training on experimental images improved performance to a precision of 0.893 ± 0.027 and a recall of 0.920 ± 0.009, and the model recovered 0.909 ± 0.010 of an independently generated CryoBlob particle list. And this CryoBlob comparison tested particle recovery rather than mask-boundary accuracy, and transfer to other particle types, ice thicknesses, contrast levels, or acquisition conditions was not evaluated.

For the quantitative tests, particle-size error was calculated only for the 2,097 of 2,250 simulated particles with matched detections, excluding missed particles and unmatched predictions. This means the reported equivalent-circular-diameter mean absolute error of 3.60 nm and mean absolute relative error of 10.1% describes size error among detected particles, which must be considered when using these methods to analyze population-size distributions. Motion correction was evaluated using simulated movies with known shifts. The implemented rigid registration corrects whole-image translation but not local or nonrigid deformation, and its low-dose performance should be validated for each specimen and acquisition setting.

Simulation-trained segmentation transferred less reliably to experimental SEM images than to experimental cryo-EM images. On the held-out SEM set the simulation-only model had low recall and precision (0.253 and 0.432, respectively). Lowering its confidence threshold increased recall only to 0.39 and increased false positives. Fine-tuning on four experimentally acquired images improved performance (recall 0.512 and precision 0.726), but still had increased false positives. These errors were not explained by insufficient pixel resolution as the spores remained 42–213 pixels across after resizing, and training at 1024 or 1280 pixels did not improve performance. Adding more than four acquired training images also did not improve performance.

Some error in this analysis comes from ambiguous reference contours. Of 1,085 annotations, 777 (72%) were confidently single spores, 190 (18%) had plausible size but ragged boundaries that could not be classified reliably from the geometry, and the remainder comprised fragments or contours that contained aggregates. Evaluating the same predictions against progressively stricter reference subsets raised recall from 0.512 for all annotations to 0.572 for plausible single objects and 0.693 for confident spores. In these cases precision remained nearly constant at 0.726, 0.714, and 0.699, respectively (Supplementary Figure S7 and Supplementary Table S8). Some false negatives corresponded to fragments, aggregates, or objects with uncertain boundaries rather than clear single spores. Removing these annotations increased recall, but this filtered value may overestimate performance because the filter can also remove genuine small or irregular spores. In this dataset, the SEM model was most dependable for isolated, high-contrast spores with clear boundaries. Debris, fragments, touching spores, and aggregates remained failure-prone or annotation-dependent cases

and should be reviewed before quantitative use.

3 Discussion

ACORN connects image segmentation and model training to measurements, provenance records, and statistical analysis in one locally executable application. EM analysis often separates these image preparation, annotation, training, inference, measurement, and statistics processes among several programs. Transfers between programs detaches masks from pixel size calibration and loses track of how annotations or measurements are produced. ACORN keeps these components intact by having segmentation and subsequent measurement and analysis methods in a shared environment. CLU provides plain-language access to the same operations available through the graphical interface, command line, and Python library while retaining each operation in the session record.

General-purpose computer-vision platforms supply annotation and model-training tools for natural images, while specialized microscopy packages perform particle picking, cellular segmentation, motion correction, or image simulation. Microscopy Image Browser, μSAM, AtomAI, and DL4MicEverywhere address complementary portions of this workflow [8, 9, 10, 11]. Our previous YOLOv11 method connected bacterial detection with morphological analysis in low-dose cryo-EM images [28], and other studies have connected particle instance segmentation to size analysis [29, 30]. These systems target particular applications and may require users to jump across software or write task-specific, one-time use code. Within ACORN the same annotation object is used for manual drawing, promptable segmentation, supervised prediction, and simulation-derived labels. Reviewed annotations keep their identifiers and modification histories during dataset construction, model training, batch inference, calibrated measurement, and statistical analysis. In this work the same software and processing steps were used for Cryo-EM nanoparticles and SEM bacterial spores datasets despite differences in contrast, scale, morphology, background, and image-formation physics.

The Cryo-EM use case showed that prompt-assisted masks could be reviewed, used to train a task-specific model, and then applied across a larger image dataset while retaining information about input image pixel size. Comparison with the independently generated CryoBlob particle list supported particle-recovery performance in these instances. The simulation TEM results addressed how well ACORN could be used to measure particle size. These results showed when using segmentation to analyze particle size that equivalent circular diameter (ECD) produced a more stable estimate than Feret diameter.

Because the simulated particles had known geometries, we could distinguish errors in the pixel-to-distance conversion (e.g., calibration) from errors caused by inaccurate segmentation boundaries. The pixel-to-distance calibration remained accurate in the software and the main discrepancies arose from differences between the predicted and true particle boundaries. Measurement error varied with both object size and measurement type, demonstrating that accurate masking is necessary for reliable particle-size estimates.

ACORN implements phase cross-correlation for inter-frame alignment and dose fractionation. Controlled simulations allowed this capability to be evaluated against known shifts and dose conditions. Alignment performance declined at the lowest tested doses, and the resulting thresholds apply only to the simulated specimens and noise conditions evaluated here. Corrected images remain linked to the acquisition and processing record as they enter reviewed detection, calibrated measurement, and statistical comparison. The goal is not to replace specialized motion-correction or single-particle-analysis packages, but to make these capabilities available within a broader quantitative imaging workflow and show they can be implemented and used for analysis in ACORN.

Simulation produced exact object labels and allows us to test segmentation ability as a function of specimen properties, acquisition conditions, dose, noise, and charging. This differs from previous studies that used simulation-only training for experimental HRTEM segmentation, simulation pretraining followed by experimental fine-tuning, and co-training with simulated and manually annotated cryo-electron tomography data [31, 32, 33]. In our SEM experiment, the effect of simulation initialization depended on both the number and identity of the experimental micrographs used for fine-tuning and showed that simulation-derived weights can provide a viable initialization for experimental fine-tuning.

The SEM results also show how image-processing and evaluation choices affect model performance. Changing the intensity-preparation method had only a modest effect, whereas changing which objects were included in the ground-truth annotations (e.g., reference definitions) produced a larger change in recall. This shows the reported metrics for segmentation depend on the model, image preparation, field composition, annotation, and reference definition. ACORN records these conditions and ties each measured object to its source image and annotation history. It also records the experimental pixel size, effect size, correction procedure, and statistical test selected by the user.

We are currently integrating ACORN with our in-house developed microscope-control software ([`https://github.com/sea-ecosystem`])to drive analysis at the instrument. Planned future work will add STEM-specific analysis. Existing tools outside the scope of this paper include a three-dimensional viewer, fluorescence-microscopy functions, and trajectory analysis for moving particles.

By connecting images with model annotations, calibrated measurements, provenance records, and statistical analysis, ACORN supports quantitative comparison of specimens and experimental conditions across low-dose cryogenic TEM, conventional TEM, and SEM. Some potential applications include evaluating nanoparticle size and morphology during drug-delivery and theranostic formulation; characterizing structural changes in biological samples; and quantifying particle distributions and interfaces in batteries, ionomers, and other soft and energy materials. ACORN provides a general workflow for converting electron microscopy data into reproducible measurements that can inform sample optimization, formulation development, and materials design.

4 Experimental Section

4.1 Software Design

ACORN uses Python with a PyQt6 interface where plugin packages can be built into the software and show up as panels in the interface. CLU maps natural-language requests to the same operations used by the graphical, command-line, and library interfaces. The reported analyses used Python 3.12.13, PyTorch 2.6.0 with CUDA 12.6, Ultralytics 8.4.41, and Ubuntu 22.04.5 on an Intel Xeon Platinum 8168 system with NVIDIA Tesla V100-SXM3-32GB GPUs.

4.2 Data Handling

ACORN accepts MRC, TIFF, HDF5, and DM4 inputs. The software recognizes multi-frame acquisitions during import and maintains them as frame stacks. Users can select a frame range and calculate either an average or a sum image. For dose-series analysis, selected frames can be divided into equal-dose bins. ACORN can then perform a two-pass rigid registration using phase cross-correlation [21, 23]. The first pass estimates the translation of each frame relative to the uncorrected stack mean. The translated frames are then averaged, and a second cross-correlation pass refines each translation relative to the aligned mean. ACORN displays each bin average and its difference from the first bin using common intensity scales. When spatial calibration is available, the software

reports the drift trajectory, displacement between successive frames, maximum displacement, and total path length in the correct units. Dose-series averages can then be calculated from the aligned frames.

The motion-and-dose simulation used for Figure 2 contains 100 frames of 512×512 pixels at 0.5 nm/pixel and 10 $e^{-}\text{Å}^{-2}$ per frame. For our case we simulated a phase-separated thin-film with a membrane-like interface and embedded nanoparticles. Local interface deformation and particle displacement relative to the interface were determined after rigid alignment. Dose-dependent contrast loss, structural redistribution, and Poisson counting noise were applied over a total dose of 1000 $e^{-}\text{Å}^{-2}$. Equal-dose averages were calculated from the aligned frames. Membrane position was estimated from the local cross-interface intensity minimum. Particle centroids were detected in the corrected dose bins and tied between bins using ACORN's distance-gated Hungarian assignment and the measurement error was calculated against the known simulated coordinates. Bin-width comparisons were centered near a cumulative dose of 500 $e^{-}\text{Å}^{-2}$ and used intervals of 20, 100, and 250 $e^{-}\text{Å}^{-2}$. The noise-free, zero-dose panel represents simulation ground truth Global-drift error was calculated as the RMSD between the recovered and programmed two-dimensional translation vectors across frames. Membrane-displacement error was the absolute difference between the recovered and programmed membrane positions. Nanoparticle-relative displacement error was calculated after subtracting the local membrane displacement from the nanoparticle displacement. Reported mean absolute errors were averaged over the corresponding tracked positions and dose bins. Nanoparticle-localization error in the bin-width comparison was the Euclidean distance between each recovered centroid and its programmed position.

Motion and dose performance was evaluated separately using 48 independently simulated movies, with six movies per dose condition and a new specimen configuration, drift trajectory, and noise realization generated for every movie. Alignment root-mean-square error was calculated by comparing the recovered trajectory with the known simulated trajectory. These results are reported in Supplementary Figure S11 and Supplementary Table S14. Image sharpness was defined as the variance of the vertical finite-difference image gradient,

$$S(I) = \mathrm{Var}(\nabla_y I)\,.$$

Sharpness gain was calculated as

$$G_S = 100\left[\frac{S(I_{\text{corrected}})}{S(I_{\text{uncorrected}})} - 1\right].$$

The independent registration results are reported in Supplementary Figure S11 and Supplementary Table S14.

Calibration-retention tests propagated known pixel sizes through metadata import, sidecar restoration, image binning, measurement, and export. We evaluated geometry consistency by comparing object dimensions calculated from exported COCO polygons with dimensions calculated from the corresponding source masks. Test definitions and results are provided in Supplementary Table S11.

4.3 Statistical analysis

We analyzed the nanoparticle and spore examples using Kruskal-Wallis omnibus tests with $\alpha = 0.05$. Post hoc pairwise comparisons used two-sided Mann–Whitney U tests with Bonferroni-adjusted p values. The omnibus p values, numbers of analyzed objects, and largest absolute Cliff's δ values are reported in the captions of Figures 3 and 4. The values of n in those figures corresponds to

numbers of objects counted in those images and these comparisons are meant to display the software output and do not support specimen-level biological conclusions. Reported training values are means and standard deviations across seeds unless otherwise stated. We calculated image-level variation using each input micrograph as the sampling unit. A micrograph containing reference objects and no correct detections was assigned F1=0. We used Spearman rank correlations for exploratory associations among dark-pixel fraction, object size, confident-reference fraction, and validation F1. Exact sample sizes and micrograph-level values are reported in Supplementary Table S15.

4.4 Provenance handling

Every annotation is assigned a stable identifier and connected to the input micrographs. The record includes origin, path, creation and modification timestamps, modification count, review status, parent and batch identifiers, spatial calibration, and statistical analysis methods. Model-generated annotations record the model name and checkpoint hash to aide in provenance tracking. To handle repeated instances of changes to an annotation, annotation geometry is serialized as sorted-key JSON after vertex coordinates are rounded to 10^{-3} pixel and then hashed using SHA-256. The geometry identifier stores the first 16 hexadecimal characters of this hash. The model-checkpoint retains the complete SHA-256 hash, file size, and fingerprints from bytes at the beginning and end of the file. During an audit the stored geometry hash is compared with a calculated hash from the current annotation geometry.

Dataset records tie the input images and annotations to their training, model-selection, and validation assignments. These also keep random seed, tiling parameters, augmentations, and negative-sampling settings. Image-processing records keep the intensity-processing method, intensity limits and stage at which the processing was applied. Each training-run record includes a dataset version, model configuration, software environment, timestamp, and resulting checkpoint hash. Prediction records identify the checkpoint used for inference. Measurement and statistical-analysis records have the contributing annotations, source images, spatial calibrations, experimental units, selected variables, exclusion criteria, statistical tests, and test parameters.

The provenance audit we ran examined the stored origins and had fields of all 1,375 final SEM annotation records across 26 source micrographs. The final store contained 1,085 imported-sidecar records, 51 SAM-origin records, and 239 YOLO-origin records. Rejected masks remained in the review log but were excluded from this denominator for our report. The measurement export contained one row per final annotation. We then selected two representative records and traced their complete analysis chains through the associated identifiers and hashes. The full audit results and the two illustrated chains are reported in Supplementary Table S12 and Supplementary Figure S9, respectively.

4.5 Simulation for TEM and SEM

TEM images are produced by multislice propagation with Fresnel transfer between slices and a weak phase contrast transfer function (including spherical aberration, amplitude contrast, and defocus) in the convention used by CTFFIND and RELION [34]. Coherence envelopes, inelastic thickness loss, dose-dependent damage envelope, parameterized detector response, and Poisson counting are all contained in the simulation. Six column configurations and nine detectors are provided in the software and more can be added by end-users.

SEM images are produced by Monte Carlo electron transport using screened Rutherford elastic scattering with the relativistic correction and a modified continuous slowing-down approximation [27]. Secondary and backscattered yields come from traced trajectories in the simulation. Interaction-

volume kernels are traced once per material, beam energy, and tilt. Kernels reproduce full trajectory quantiles to better than 1% (Supplementary Figure S1). Seventeen materials and six scenes are included by default with ACORN install, but others can be added by end users. Supplementary Tables S1 and S2 report the full parameter sets, Supplementary Table S3 reports the specimen models, and Supplementary Table S4 reports the generated datasets.

4.6 Forward-model benchmarking

Published backscattered yields and electron-range values were used to benchmark selected outputs of the SEM forward model [26, 27, 35]. TEM implementation consistency was evaluated by recovering specified defocus values from simulated image power spectra. These comparisons test the implemented models under the stated conditions but do not establish general physical accuracy. Accepted benchmark outputs were encoded as regression tests so that subsequent software changes that alter the results can be detected.

4.7 Dataset preparation and training

We used ACORN to produce SAM-3 generated masks for the SEM and cryo-EM imgaes. These annotations were then reviewed and verified by experts. Regions were accepted based on visible boundary of the target structure. Accepted annotations were used to construct model-training datasets. During dataset finalization, ACORN extracted image tiles, applied augmentations, sampled negative regions, and generated reproducible data partitions using a recorded random seed. Source micrographs were assigned to the training, model-selection, and validation partitions prior to tiling. Micrographs used during prompt development are in the training partition.

For simulation test evaluation, we applied the final checkpoint selected on each model-selection partition once to the disjoint 60-image test partition without additional tuning. Reported precision, recall, F1, and mean average precision values are for the target class rather than macro averages over target and debris classes. For measurement evaluation, predicted and ground-truth instances were assigned one-to-one in descending mask intersection-over-union, requiring matching class and an intersection-over-union of at least 0.5. Bias, mean absolute error, and mean absolute relative error were calculated among matched detections, and source images were used as the resampling unit.

Detection and segmentation models were trained within ACORN (Supplementary Tables S5 and S10; Supplementary Figure S3). Recorded training settings include the model implementation and version, pretrained checkpoint, input normalization, image size, optimizer, learning rate, number of epochs, batch size, augmentations, negative-sampling settings, and random seed. Datasets can be exported in COCO format with run-length and polygon representations for each instance, in detection-label format, and in NeXus format with calibrated axes. Models and associated metadata can be prepared through a public model hub.

The experimental TEM and SEM models were trained through ACORN using three seeds per condition. For the SEM intensity-preparation the annotations, input micrograph partitions, hyperparameters, and training seeds were held constant. Three pre-export intensity preparations were compared: first the original per-image 0.5-99.5% stretch; second, the original 8-bit SEM intensities; and third, a single global intensity window applied across all 26 images. ACORN's training exporter subsequently applied its standard percentile normalization before tiling in all three conditions.

For the simulation-initialization experiment, training micrographs were resampled from a pool of eight for the two- and four-micrograph conditions and performance was evaluated on 18 separate

validation micrographs. This legacy partition was distinct from the later 26-micrograph SEM benchmark. Four independently resampled training sets were evaluated for each initialization strategy and micrograph count. For the eight-micrograph condition, the complete pool was used and ten recorded fine-tuning seeds were evaluated for each initialization strategy. All simulation-initialized conditions reused one simulation-pretrained checkpoint and therefore were not independent simulation-pretraining replicates. Under the configuration recorded as `deterministic=True`, all simulation-initialized models produced identical weights and predictions. The natural-image-initialized models produced distinct weights.

4.8 Spore processing

Bacterial spores used for SEM imaging were produced by nutrient exhaustion. *Bacillus thuringiensis* strains were sporulated in liquid modified G medium [36] containing yeast extract (2 g,L^{-1}), $(NH_4)_2SO_4$ (2 g,L^{-1}), $CaCl_2 \cdot 2H_2O$ (25 mg,L^{-1}), K_2HPO_4 (0.5 g,L^{-1}), $MgSO_4 \cdot 7H_2O$ (0.2 g,L^{-1}), $MnSO_4 \cdot 4H_2O$ (50 mg,L^{-1}), $ZnSO_4 \cdot 7H_2O$ (5 mg,L^{-1}), $CuSO_4 \cdot 5H_2O$ (5 mg,L^{-1}), and $FeSO_4 \cdot 7H_2O$ (0.5 mg,L^{-1}), adjusted to pH 7.1. *B. subtilis*, *B. spizizenii*, *B. vallismortis*, *B. mojavensis*, and *B. inaquosorum* strains were sporulated in liquid Difco sporulation medium (DSM) [37].

Strains were grown overnight in 5 mL of lysogeny broth containing casein peptone (10 g,L^{-1}), NaCl (10 g,L^{-1}), and yeast extract (5 g,L^{-1}) in a shaking incubator at 37 °C and 250 rpm. Cells were washed with 1× phosphate-buffered saline (pH 7.4). The resulting pellets were resuspended in either modified G medium or DSM, and the cultures were incubated at 37 °C for 72 h.

Spores were harvested by centrifugation at 1,000 × g for 15 min at room temperature and resuspended in 350 μL of 20% Histodenz. Each spore suspension was carefully layered over 1.5 mL of 50% Histodenz in a microcentrifuge tube and centrifuged at 16,000 × g for 30 min at room temperature. The resulting pellets were washed five times with sterile Milli-Q water. Purified spores were resuspended in Milli-Q water and stored at 4 °C until analysis.

4.9 Spore SEM data collection

SEM images were acquired using a Zeiss Merlin scanning electron microscope operated at an accelerating voltage of 1 kV and a beam current of 100 pA. Images were collected using either the in-lens or SE2 detector with Zeiss SmartSEM software (version 6.06). A 50 μL aliquot of the prepared spore suspension was deposited onto carbon tape mounted on an aluminum pin stub. The sample was loaded into the microscope and dried during the initial chamber pump-down before imaging.

4.10 Nanoparticle synthesis

PLGA (Resomer RG 503 H; poly(D,L-lactide-co-glycolide), 50:50 lactide:glycolide, acid-terminated, MW = 24,000–38,000, and the designated formulation components were dissolved or dispersed in acetone solutions. An aqueous solvent solution containing vitamin E-TPGS (D-alpha-tocopherol polyethylene glycol 1000 succinate was prepared (0.5% wt in 18.2 M$\Omega \cdot$ cm deionized water) as collection solution. The organic solutions of PLGA (40 mg/mL) in acetone and payload formulation, described below, were rapidly combined in a confined impinging jet mixer, and the effluent was collected in the aqueous TPGS solution. The resulting suspensions were reacted at 37 °C for 3 days to allow particles to harden, then solvent was removed by evaporation. Nanoparticles were collected by centrifugation (15,000 rpm, 20 min, 8 °C), washed at three times with deionized water, and resuspended in 1 mL deionized water after each wash. Three formulations were prepared: (1) control PLGA nanoparticles; (2) nanoparticles treated with La(NO3)3 in NH4OAc, at a La concentration

of 5.8 mM; and (3) nanoparticles treated with DOTA and La(NO3)3, at nominal concentrations of 0.17 mM DOTA and 5.8 mM La.

4.11 Nanoparticle cryo-EM preparation

Nanoparticle samples were prepared on Quantifoil R2/2, 300-mesh copper grids. Before sample application, the grids were glow discharged for 45 s at 5 mA using a GloQube glow-discharge system. Vitrification was performed using a Vitrobot with the chamber maintained at 4 °C and 100% relative humidity. A 5 μL aliquot of sample was applied to each grid. After a 4 s wait time, the grid was blotted once for 5 s using a blot force of −2 and then vitrified in liquid nitrogen cooled liquid ethane.

4.12 Nanoparticle cryo-EM data collection

Nanoparticle cryo-EM data were acquired on a ThermoFisher Titan Krios G4 operated at 300 kV in energy-filtered TEM (EFTEM) nanoprobe mode. Imaging used a 70 μm C2 aperture, spot size 9, and no objective aperture. Images were recorded on a Gatan K3 direct electron detector at nominal magnifications ranging from 8,700× to 61,000×. The total accumulated electron dose was approximately 60 e^{-}Å^{-2}.

5 Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science FWP ERKCZ64, Structure Guided Design of Materials to Optimize the Abiotic-Biotic Material Interface, as part of the Biopreparedness Research Virtual Environment (BRaVE) initiative. For this research sample preparation, electron microscopy (cryo-EM and SEM), and image-analysis work was conducted as part of multiple user projects at the Center for Nanophase Materials Sciences (CNMS), a U.S. Department of Energy Office of Science User Facility at Oak Ridge National Laboratory. Cryo-EM data were collected using instrumentation within ORNL's Materials Characterization Core provided by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the DOE. SEM spore material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Biological Systems Science Division, through the Critical Minerals and Materials Pilot Program under project ERKPA87. The nanoparticle work was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy.

6 Author contributions

Alexis N. Williams: Conceptualization, Methodology, Software Development, Cryo-EM data collection, Validation, Formal analysis, Data curation, Visualization, Writing: original draft, Writing: review and editing, Supervision, Project administration.
Chanda R. Harris: Cryo-EM sample preparation. Writing: review and editing.
Albina Y. Borisevich: EM data collection and analysis. Writing: methods, review, and editing.
Michael J. Zachman: Software development. Writing: review and editing.
Juhyun Oh: Software development. Writing: review and editing.
Eric R. Hoglund: Software development. Writing: review and editing.
Asraful Haque: Editing of code and software development. Writing: review and editing.
Sumner B. Harris: Editing of code and software development. Writing: review and editing.
Jack C. Lasseter: SEM data collecting. Writing:review and editing.

Leah H. Hochanadel: Spore preparation and SEM analysis. Writing: methods, review, and editing.
Delyana P. Vasileva: Spore preparation and analysis. Writing: methods, review, and editing.
Miguel Toro-Gonzále: Nanoparticle production. Writing: review
Jennifer L. Morrell-Falvey: Methodology and supervision. Writing:review and editing
Josh K. Michener: Writing:review and editing Supervision, Project administration.
Sandra M. Davern: Writing:review and editing Supervision, Project administration.
Brian C. Sanders: Writing:review and editing Supervision, Project administration.
Scott T. Retterer: Writing:review and editing Supervision, Project administration.

7 Supporting information

The Supporting Information contains simulation settings and benchmarks, model-training configurations, training curves, simulation test and simulation-to-experimental transfer results, validation predictions, reference-set composition analyses, software comparisons, calibration-retention tests, provenance audits, measurement-accuracy results, and SEM intensity-preparation analyses (Supplementary Figures S1–S10 and Supplementary Tables S1–S16).

8 Dataset and code availability

The ACORN source code and versioned reproducibility package supporting this article are publicly available at `https://github.com/vitrifyai/acorn_software` and `https://github.com/vitrifyai/acorn-measurements-paper`. The GitHub package contains numerical source data underlying the figures and tables, image and annotation identifiers, dataset-partition manifests, evaluation records, model-checkpoint hashes, training and simulation configurations, random seeds, motion-correction trajectories, dose-response data, selected example images, and scripts for reproducing the reported figures, metrics, and statistical analyses.

The complete archival dataset, including native microscopy files, representative raw cryo-EM and SEM data, derived and simulation datasets, reference annotations, trained model checkpoints, analysis records, final figures, and manuscript materials, is available through the Oak Ridge Leadership Computing Facility Constellation repository at `doi:10.13139/ORNLNCCS/3920044`. File-level SHA-256 manifests are provided with both packages to support integrity and provenance verification.

ACORN-owned software is distributed under the MIT License. Data, images, annotations, measurements, figures, and other publication materials are distributed under the Creative Commons Attribution 4.0 International License unless otherwise identified. Model checkpoints trained using Ultralytics YOLO are distributed under the GNU Affero General Public License version 3 and remain subject to applicable Ultralytics upstream licensing terms.

9 LLM use

Claude was used to assist with grammatical editing, software formatting/coding issues, and LaTeX organization/scripting errors.

References

[1] Roboflow. Roboflow: Computer vision dataset management, annotation, and deployment. Accessed 2026-09-09.

[2] CVAT.ai. Computer vision annotation tool. Accessed 2026-09-09.

[3] HumanSignal. Label studio. Accessed 2026-09-09.

[4] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, and B. Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1153–1160, 2019.

[5] T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum, D. Quentin, D. Roderer, S. Tacke, B. Siebolds, E. Schubert, T. R. Shaikh, P. Lill, C. Gatsogiannis, and S. Raunser. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-em. *Communications Biology*, 2:218, 2019.

[6] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu. Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021.

[7] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 265–273. Springer, 2018.

[8] Ilya Belevich, Merja Joensuu, Darshan Kumar, Helena Vihinen, and Eija Jokitalo. Microscopy image browser: A platform for segmentation and analysis of multidimensional datasets. *PLOS Biology*, 14(1):e1002340, 2016.

[9] Anwai Archit, Lennart Freckmann, Sibi Nair, et al. Segment anything for microscopy. *Nature Methods*, 22:579–591, 2025.

[10] Maxim Ziatdinov, Ayana Ghosh, Chun Yin Wong, and Sergei V. Kalinin. AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nature Machine Intelligence*, 4:1101–1112, 2022.

[11] Iván Hidalgo-Cenalmor, Joanna W. Pylvänäinen, Mariana G. Ferreira, et al. DL4MicEverywhere: Deep learning for microscopy made flexible, shareable and reproducible. *Nature Methods*, 21:925–927, 2024.

[12] T. Verduin, S. R. Lokhorst, and C. W. Hagen. Nebula: Monte carlo simulator of electron-matter interaction. *SoftwareX*, 12:100605, 2020.

[13] J. S. Villarrubia et al. Scanning electron microscope measurement of width and shape of 10 nm patterned lines using a JMONSEL-modeled library. *Ultramicroscopy*, 154:15–28, 2015.

[14] D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin. CASINO V2.42: A fast and easy-to-use modeling tool for scanning electron microscopy and microanalysis users. *Scanning*, 29(3):92–101, 2007.

[15] K. T. Arat, T. Klimpel, and C. W. Hagen. Model sensitivity analysis of monte-carlo based SEM simulations. *Results in Physics*, 19:103545, 2020.

[16] I. Lobato, S. Van Aert, and J. Verbeeck. Progress and new advances in simulating electron microscopy datasets using MULTEM. *Ultramicroscopy*, 168:17–27, 2016.

[17] C. Ophus. A fast image simulation algorithm for scanning transmission electron microscopy. *Advanced Structural and Chemical Imaging*, 3:13, 2017.

[18] J. Barthel. Dr. probe: A software for high-resolution STEM image simulation. *Ultramicroscopy*, 193:1–11, 2018.

[19] J. Madsen and T. Susi. The abTEM code: Transmission electron microscopy from first principles. *Open Research Europe*, 1:24, 2021.

[20] E. J. Kirkland. *Advanced Computing in Electron Microscopy*. Springer, Cham, 3 edition, 2020.

[21] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R. Booth, Michael B. Braunfeld, Sander Gubbens, David A. Agard, and Yifan Cheng. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em. *Nature Methods*, 10(6):584–590, 2013.

[22] Timothy Grant and Nikolaus Grigorieff. Measuring the optimal exposure for single-particle cryo-em using a 2.6 å reconstruction of rotavirus VP6. *eLife*, 4:e06980, 2015.

[23] Shawn Q. Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332, 2017.

[24] Manuel Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup. Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158, 2008.

[25] Debangshu Mukherjee, Alexis N. Williams, Spenser R. Cox, Marshall T. McDonell, and Albina Y. Borisevich. cryoblob. Software, 2025.

[26] L. Reimer. *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Springer, Berlin, 2 edition, 1998.

[27] D. C. Joy. *Monte Carlo Modeling for Electron Microscopy and Microanalysis*. Oxford University Press, New York, 1995.

[28] Sita Sirisha Madugula, Lynnicia N. Massenburg, Spenser R. Brown, Amber N. Bible, Chanda R. Harris, Lance X. Zhang, Kiara Parker, Scott T. Retterer, Jennifer L. Morrell-Falvey, Rama K. Vasudevan, and Alexis N. Williams. Automated bacterial identification and morphological feature analysis in low-dose cryo-EM using YOLOv11. *Advanced Intelligent Discovery*, 2026.

[29] Bastian Rühle, Julian Frederic Krumrey, and Vasile-Dan Hodoroaba. Workflow towards automated segmentation of agglomerated, non-spherical particles from electron microscopy images using artificial neural networks. *Scientific Reports*, 11:4942, 2021.

[30] Batuhan Yildirim and Jacqueline M. Cole. Bayesian particle instance segmentation for electron microscopy image quantification. *Journal of Chemical Information and Modeling*, 61(3):1136–1149, 2021.

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

[32] Shengmin Zhou, Xuehai Huang, Junjie Lin, and Yu Wang. A physics-informed synthetic-to-experimental framework for few-shot structural segmentation of HRTEM images. *npj Computational Materials*, 2026. Advance online publication.

[33] Linh T. Nguyen, Carson Purnell, and Matt Swulius. Co-training generalized segmentation networks with real and synthetic data. *Microscopy and Microanalysis*, 31(Supplement 1):ozaf048.477, 2025. Conference abstract.

[34] A. Rohou and N. Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015.

[35] K. Kanaya and S. Okayama. Penetration and energy-loss theory of electrons in solid targets. *Journal of Physics D: Applied Physics*, 5(1):43–58, 1972.

[36] H. U. Kim and J. M. Goepfert. A sporulation medium for Bacillus anthracis. *Journal of Applied Bacteriology*, 37(2):265–267, 1974.

[37] Pierre Schaeffer, Jacques Millet, and Jean-Paul Aubert. Catabolic repression of bacterial sporulation. *Proceedings of the National Academy of Sciences of the United States of America*, 54(3):704–711, 1965.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AcornACS17.pdf