Supporting Information

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

Alexis N. Williams[1,*], Chanda R. Harris[1], Albina Y. Borisevich[1], Michael J. Zachman[1], Juhyun Oh[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], Josh 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, United States
[2]Radioisotope Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States
[3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States
[4]NorthStar Medical Technologies Beloit, Wisconsin 53511, USA

[*]Corresponding author: williamsan@ornl.gov

Supplementary Table S1: TEM forward-model settings and benchmarking. CTF denotes contrast transfer function; DQE, detective quantum efficiency; MTF, modulation transfer function; and MAE, mean absolute error.

Parameter	**Value**
Forward model	Multislice propagation through vitreous ice
Transfer function	Weak-phase CTF using the CTFFIND/RELION convention
$\chi(k)$	$\pi\lambda\,\Delta f\,k^2 - \frac{1}{2}\pi C_{\mathrm{s}}\lambda^3 k^4$
Envelopes	temporal, spatial, B-factor, dose damage
Microscopes	Krios, Krios-CFEG, Glacios, Talos Arctica, Talos L120C, and C_{s}-corrected
Detectors	K3, K2, Falcon4, Falcon4i, Falcon3EC, Apollo, DE64, Ceta, ideal
Detector response	Parameterized DQE(q) and MTF(q); representative response shapes, not vendor curves
Slice thickness	40 Å
Inelastic loss	Thickness dependent
Benchmark	Defocus recovered from simulated power spectra: $R^2 = 0.986$ and MAE $= 0.163\,\mu$m

Supplementary Table S9: **Representative software platforms and studies with capabilities relevant to the ACORN workflow.** The table summarizes functionality documented in published software. It is intended to identify functional overlap and differences in scope for ACORN. Capabilities available only through external plug-ins or separately configured components are described accordingly.

Platform or study	**Primary scope**	**Capabilities overlapping ACORN**	**Distinction relative to ACORN**
General annotation and computer-vision platforms			

Supplementary Table S9 continued

Platform or study	Primary scope	Capabilities overlapping ACORN	Distinction relative to ACORN
CVAT [1]	Image and video annotation	Manual annotation, model-assisted annotation, review workflows, dataset export, command-line utilities, and local deployment.	Does not natively provide EM metadata interpretation, spatial-calibration propagation, calibrated object measurement, EM simulation, or a unified link from trained-model predictions to statistical results.
Label Studio [2]	Configurable data annotation	Manual and model-assisted annotation, review, machine-learning backends, dataset export, APIs, and local deployment.	Machine-learning integration depends on separately configured backends. EM calibration, physical measurement, image simulation, and microscopy-specific analysis provenance are outside its primary scope.
napari [3]	Extensible multi-dimensional image viewing	Interactive visualization, label editing, measurement, scripting, and access to segmentation and machine-learning methods through plug-ins.	Functions depend on the installed plug-ins and generally do not share a single training-to-measurement provenance model. Native EM metadata handling and statistical reporting are not its primary focus.
General and microscopy-specific image-analysis platforms			
Fiji/ImageJ [4, 5]	General scientific image processing	Microscopy-format import through Bio-Formats, spatial calibration, manual regions of interest, image processing, measurement, plotting, scripting, batch analysis, and machine-learning functions through plug-ins.	Provides a broad plug-in ecosystem rather than one structured record connecting annotations, training datasets, model checkpoints, predictions, measurements, and statistical outputs.

Supplementary Table S9 continued

Platform or study	Primary scope	Capabilities overlapping ACORN	Distinction relative to ACORN
Microscopy Image Browser [6]	Segmentation and analysis of multi-dimensional microscopy data	Microscopy-format import, voxel-dimension tracking, manual and machine-learning-assisted segmentation, quantitative morphology, result export, plug-ins, and processing logs.	Primarily supports interactive segmentation and quantification. ACORN additionally records annotation origin and revision history and connects task-specific deep-model training, checkpoint identity, batch predictions, measurements, and statistical records.
Cellpose [7]	Generalist biological-image segmentation	Graphical, command-line, and Python interfaces; interactive correction; custom model training; batch inference; and two- and three-dimensional segmentation.	Focused on biological-object segmentation. It does not natively manage EM calibration, simulation, downstream statistical analysis, or annotation-to-checkpoint-to-measurement provenance.
Segment Anything for Microscopy (μSAM) [8]	Promptable microscopy segmentation and model adaptation	Point- and box-prompted segmentation, automated segmentation, high-throughput annotation, EM-specific models, multidimensional segmentation, tracking, and in-application SAM fine-tuning.	Focused on SAM-based annotation and model adaptation. ACORN also supports text-prompted segmentation and uses reviewed annotations to train smaller task-specific models while retaining calibration, measurements, statistical outputs, and analysis provenance.
Electron-microscopy and structural-biology platforms			
RELION [9, 10]	Single-particle cryo-EM and cryo-ET processing	Native cryo-EM metadata, particle selection, motion correction, contrast-transfer-function estimation, classification, reconstruction, workflow execution, and local or cluster-based processing.	Designed for structure determination rather than general EM annotation, physical object measurement, statistical comparison of specimen groups, or editable annotation provenance across multiple EM modalities.

Supplementary Table S9 continued

Platform or study	Primary scope	Capabilities overlapping ACORN	Distinction relative to ACORN
cryoSPARC [11]	Single-particle cryo-EM and cryo-ET processing	Cryo-EM metadata handling, motion correction, particle picking, classification, reconstruction, workflow management, and local or cluster execution.	Optimized for structural-biology reconstruction workflows. General morphological measurement, user-editable annotation histories, and experiment-level statistical analysis are outside its primary scope.
AtomAI [12]	Deep learning for electron and scanning-probe microscopy	Deep-learning segmentation, model training and inference, quantitative descriptors, statistical and graph analyses, and connections between microscopy and simulation.	Primarily a Python framework for atomic, mesoscopic, and spectroscopic analyses. ACORN provides an interactive review workflow and explicit links among calibrated images, annotation revisions, dataset partitions, checkpoints, object measurements, and statistical outputs.
Portable model-training and reproducibility platforms			
DL4Mic Everywhere [13]	Portable deep-learning workflows for microscopy	Low-code model training and inference, graphical and headless execution, local and high-performance computing, containerized environments, version tracking, and multiple segmentation methods.	Emphasizes reproducible execution and sharing of deep-learning methods. ACORN additionally manages EM images, spatial calibration, editable annotations, physical measurements, and statistical results within the same application data model.
Synthetic-data and simulation-assisted EM segmentation			

Supplementary Table S9 continued

Platform or study	Primary scope	Capabilities overlapping ACORN	Distinction relative to ACORN
EMcopilot [14]	Generated training data for electron-micrograph segmentation	SAM-derived morphological information, synthetic EM image generation, modeling of noise and imaging artifacts, segmentation-model training, and nanoparticle-size analysis.	Uses learned mask-to-image generation and domain adaptation. ACORN uses parameterized TEM and SEM forward models and manages synthetic, imported, prompt-generated, and trained-model annotations within the same review, measurement, and provenance workflow.
Physics-informed synthetic-to-experimental HRTEM segmentation [15]	Few-shot multiclass HRTEM segmentation	Physics-informed image generation with co-registered structural labels, synthetic pretraining, experimental fine-tuning, and evaluation of synthetic-to-experimental transfer.	Targets multiclass structural segmentation in HRTEM. ACORN uses simulation as one component of a broader workflow spanning annotation, task-specific training, batch inference, calibrated object measurement, statistics, and provenance.
Integrated ACORN workflow			
ACORN, this work	Calibrated and traceable quantitative EM analysis	Native EM data import; calibration tracking; manual, prompt-assisted, imported, and model-generated annotations; annotation review; task-specific model training; batch inference; motion correction; TEM and SEM simulation; calibrated measurements; statistical analysis; graphical, command-line, Python, and natural-language interfaces; and local execution.	Uses a common editable annotation record to connect source images and calibration with annotation history, dataset partitions, model checkpoints, predictions, physical measurements, and statistical outputs.

Supplementary Table S2: SEM forward-model settings and benchmarking. ETD denotes Everhart–Thornley detector; TLD, through-lens detector; BSE, backscattered electron; and CDF, cumulative distribution function.

Parameter	**Value**
Forward model	Monte Carlo electron transport
Elastic scattering	Screened Rutherford model
Stopping power	Joy–Luo modified Bethe model
Range	Kanaya–Okayama model (1972)
Materials	17, including 6 compounds represented by weight-averaged Z and A
Free parameter	ε, energy deposited per escaping secondary electron; fitted by material
Fitted ε range	57–243 eV
Emission	SE1/SE2 split with material-specific interaction-volume kernels
Kernel construction	Enclosed-weight CDF derived from traced trajectories
Kernel fidelity	Less than 1% deviation for tested trajectory quantiles (Supplementary Figure S1B)
Detectors	ETD, TLD, and BSE
Optional artifact	Specimen charging; disabled by default
Benchmark	Backscatter yields within 15% of reference values for the tested materials (main-text Table 1)

Supplementary Table S3: Specimen and acquisition parameters sampled to generate the synthetic TEM and SEM datasets.

Specimen model	**Parameter varied**	**Range**
PLGA nanoparticles	Diameter	14–210 nm; log-normal, centered near 42 nm
	Particles per micrograph	15–60
	Ice thickness	40–90 nm
	Defocus	-1.0 to -3.0 μm
	detector	K3, K2, Falcon4
	Total dose	20, 40, or 60 $\mathrm{e^-\ \AA^{-2}}$
	Pixel size	15, 18, or 22 Å
	Debris	0 or 6–30 chains; chain length 70–180 nm
Bacterial spores	Length	800–3000 nm; log-normal, centered near 1267 nm
	Aspect ratio	1.05–3.00; mean 1.65
	Areal density	0.35–0.95 $\mu\mathrm{m}^{-2}$
	Coating	0 nm or 8–16 nm Au
	Beam energy	0.8–1.5 kV uncoated; 3–6 kV coated
	Charging strength	0, 0.4, 0.8, or 1.2
	Substrate	Resin, carbon, biological material, or silicon
	Pixel size	4.0, 6.5, 9.6, 14.0, or 19.3 nm
	Debris	0–13 clumps; grain diameter 45–110 nm

Supplementary Table S4: Composition and partitions of the synthetic datasets. Individual instances were defined by the simulated micrographs that provides the exact masks without manual annotation. Training, model-selection, and test images were generated from disjoint simulation-seed ranges. Final checkpoints were selected on the model-selection partition and evaluated once on the test partition without additional tuning.

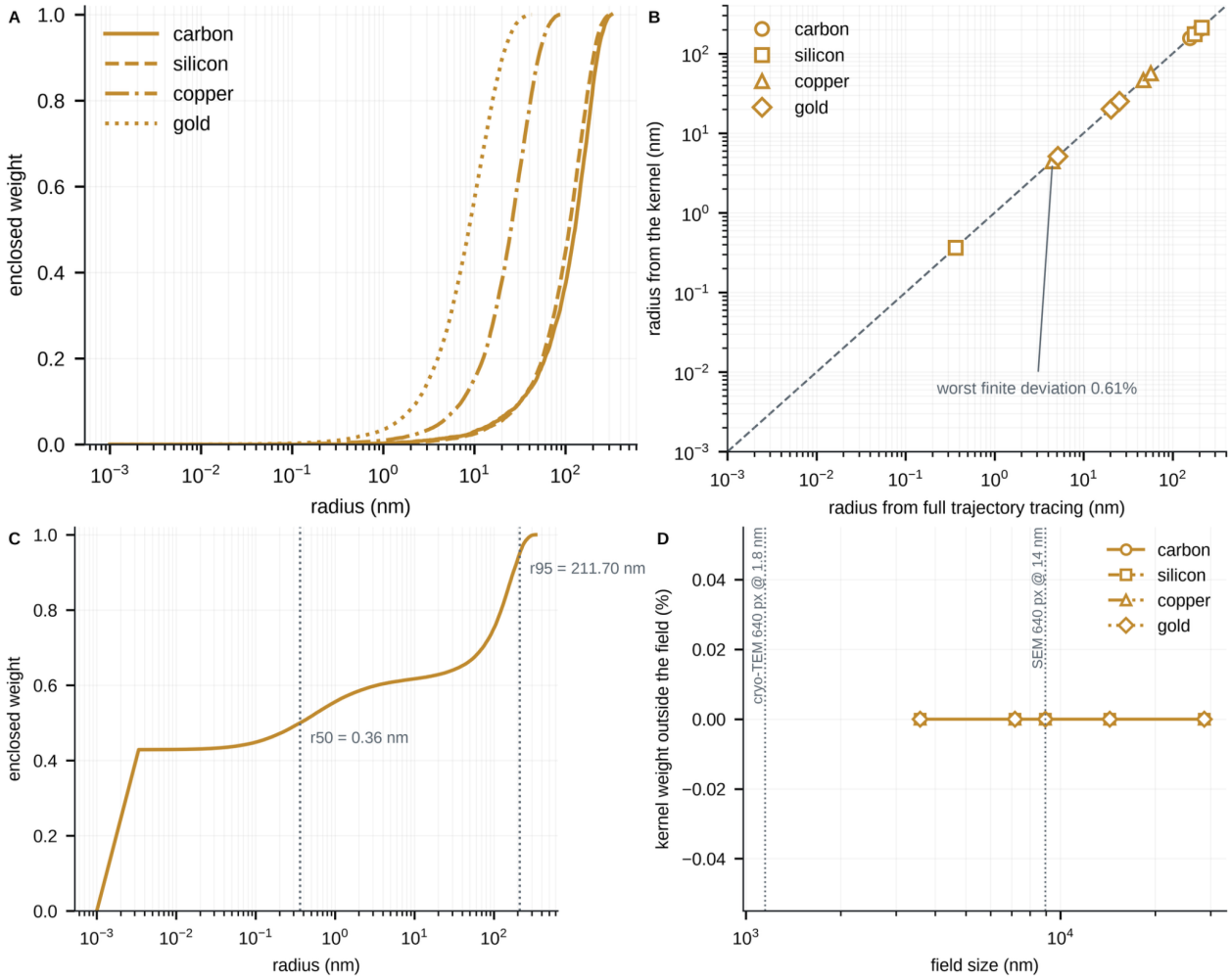
Dataset	**Images**	**Training**	**Model selection**	**Test**	**Labeled instances**	**Test fraction**
Transmission, PLGA	400	280	60	60	18,354	0.15
Scanning, spores	400	280	60	60	7,629	0.15
Total	800	560	120	120	25,983	0.15

Supplementary Table S5: Training configuration used for the synthetic-data experiments. Unless otherwise stated, both modalities used the same settings. Fine-tuning on acquired images used a batch size of 4.

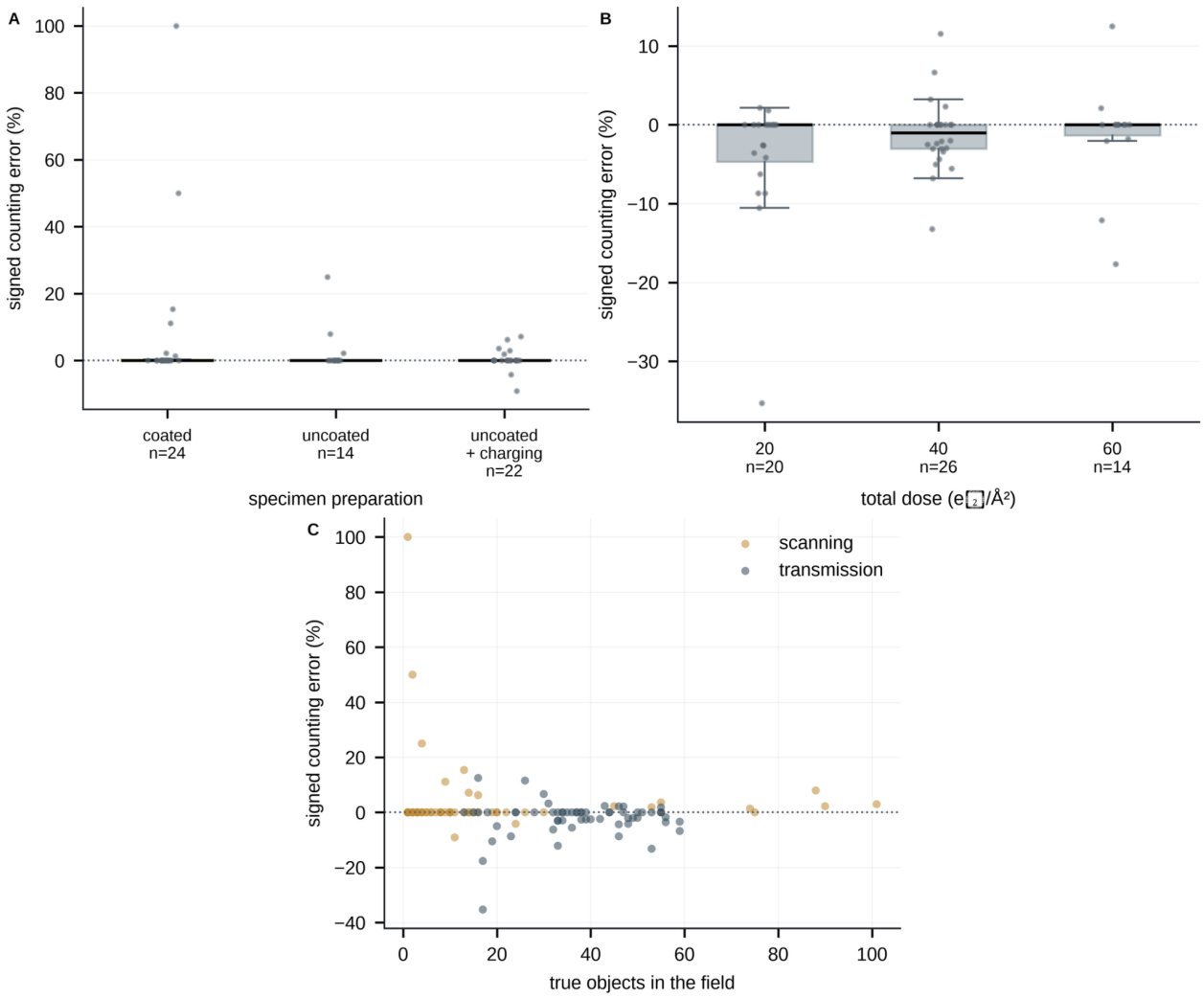
Setting	**Value**
Architecture	YOLO11s-seg
Initialization	Natural-image pretrained or simulation initialized, as stated
Epochs requested	100
Early stopping patience	25
Batch size	16 (4 when fine-tuning on acquired images)
Image size	640 px
Augmentation	Ultralytics defaults
Training/model-selection/test split	280/60/60 images generated from disjoint simulation-seed ranges
Seed	0, deterministic
Hardware	one Tesla V100-SXM3-32GB
Inference time	Approximately 7 ms per 640×640-pixel image

Supplementary Table S6: Target-class performance on held-out synthetic test images. Final checkpoints were selected on the model-selection partitions and evaluated once on disjoint 60-image test partitions without additional tuning. Precision (P), recall (R), F1, and mean average precision (mAP) are reported for the target class rather than as macro averages over target and debris classes. The 50–95 metric is averaged over intersection-over-union thresholds from 0.50 to 0.95.

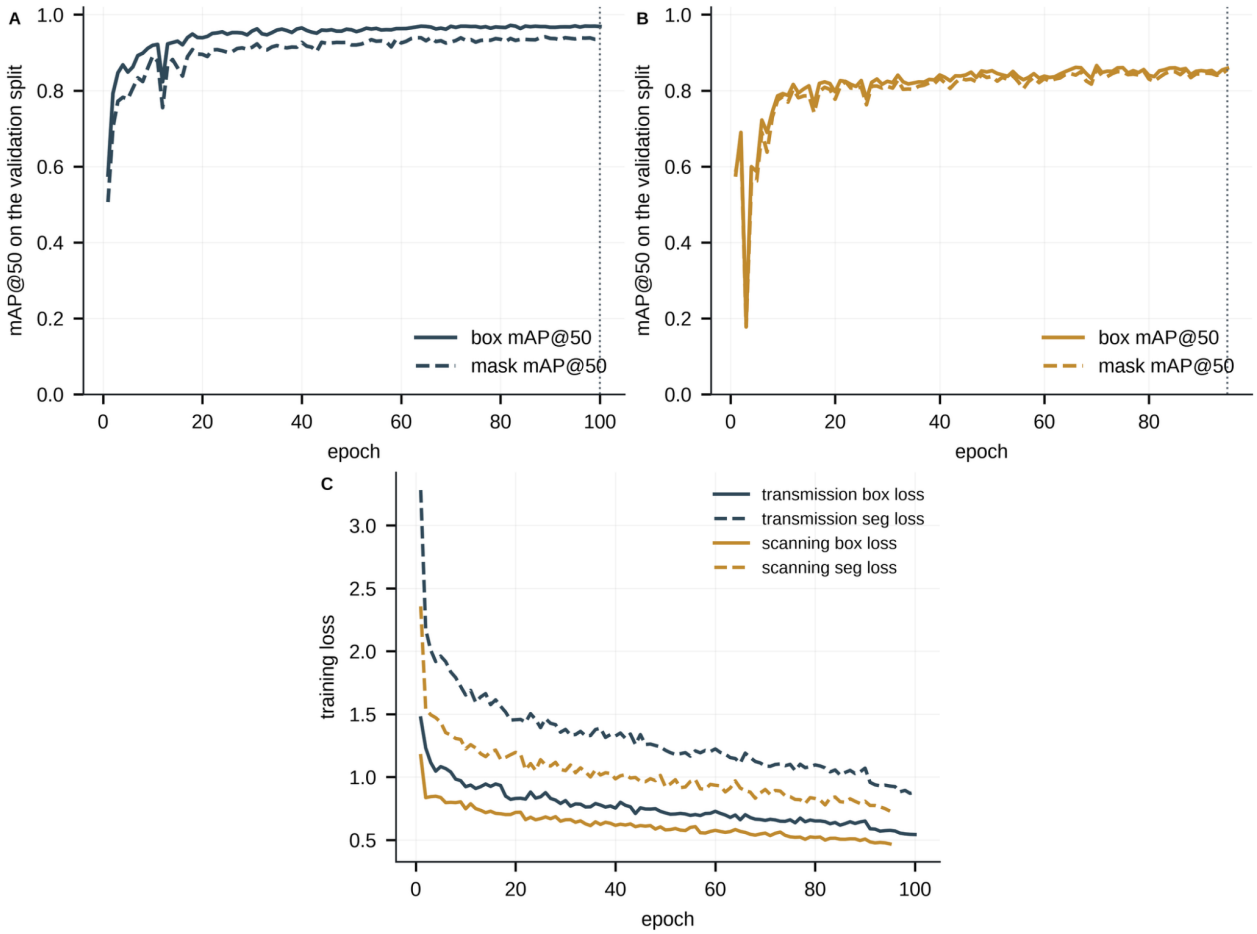
Dataset	**Images**	**Targets**	**P**	**R**	**F1**	**Box mAP$_{50}$**	**Mask mAP$_{50}$**	**Mask mAP$_{50–95}$**
Transmission, PLGA	60	2,250	0.944	0.932	0.938	0.984	0.935	0.577
Scanning, spores	60	1,083	0.972	0.988	0.980	0.974	0.990	0.872



Supplementary Figure S1: **Fidelity of the SEM interaction-volume kernels.** Monte Carlo trajectories were generated at 5 keV using 40,000 incident electrons per material. (A) Enclosed kernel weight as a function of radius for carbon, silicon, copper, and gold. These colors identify materials and line styles identify kernel summaries. (B) Kernel radii plotted against the corresponding quantiles from full trajectory tracing. The dashed line denotes equality; the largest finite deviation was 0.61%. (C) Enclosed-weight curve for the silicon secondary-electron kernel, with the median (r_{50}) and 95th-percentile (r_{95}) radii marked. (D) Fraction of kernel weight outside the simulated micrograph as a function of micrograph size. Vertical lines mark the two micrograph sizes used in this study. Across the sampled materials, energy, and micrograph sizes, the largest finite kernel-radius deviation was 0.61% and no kernel weight was lost.



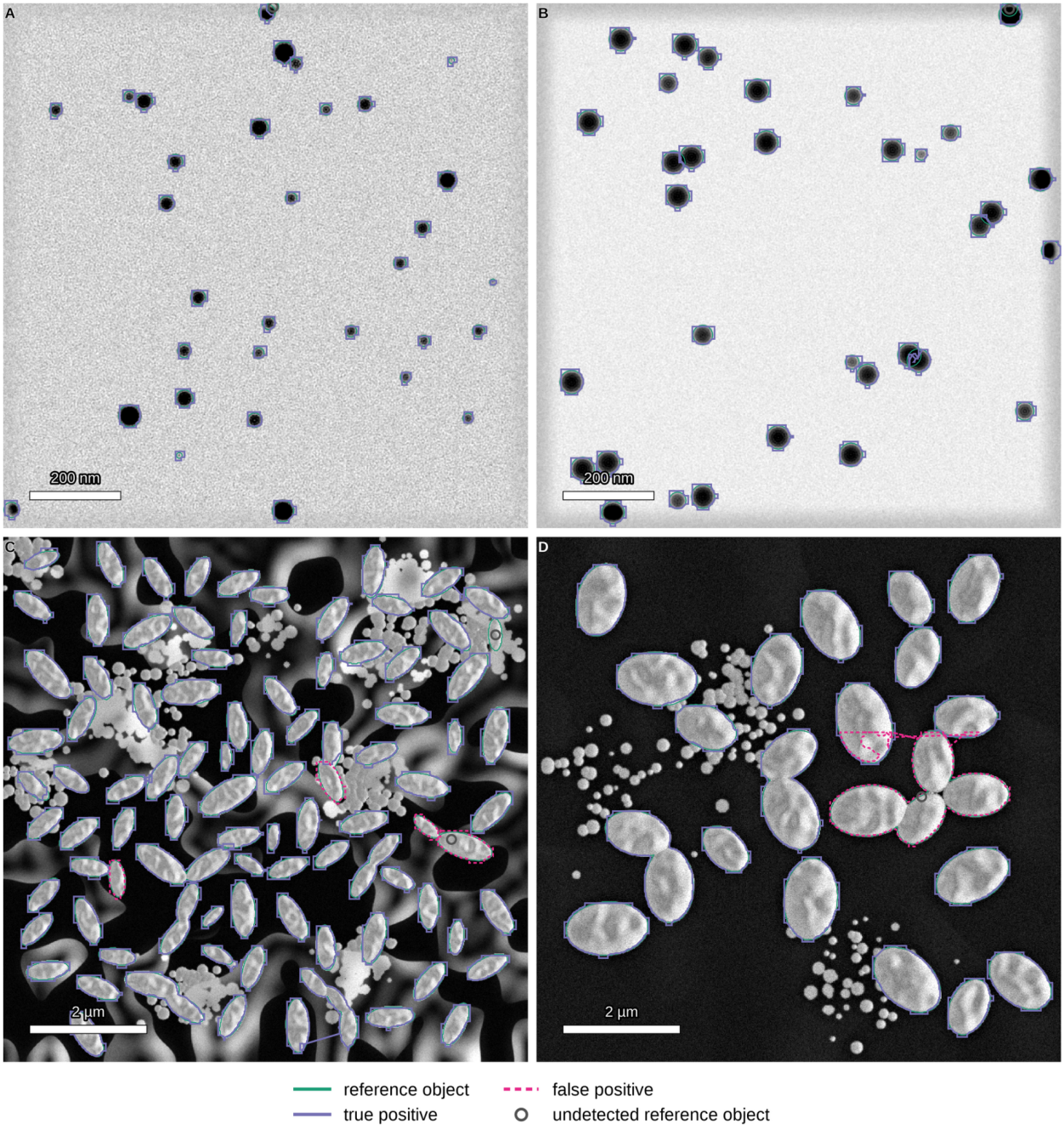
Supplementary Figure S2: **Condition-specific counting errors on held-out synthetic test images.** Signed counting error was calculated as $(N_{\text{predicted}} - N_{\text{true}})/N_{\text{true}} \times 100\%$. (A) SEM specimen conditions: coated ($n = 24$), uncoated ($n = 14$), and uncoated with charging ($n = 22$). (B) TEM total-dose conditions: 20 $\text{e}^- \text{Å}^{-2}$ ($n = 20$), 40 $\text{e}^- \text{Å}^{-2}$ ($n = 26$), and 60 $\text{e}^- \text{Å}^{-2}$ ($n = 14$). Boxes show the interquartile range, center lines show medians, and whiskers extend to 1.5 times the interquartile range. The points show individual micrographs. (C) Counting error against the true number of objects per micrograph for the SEM and TEM datasets.



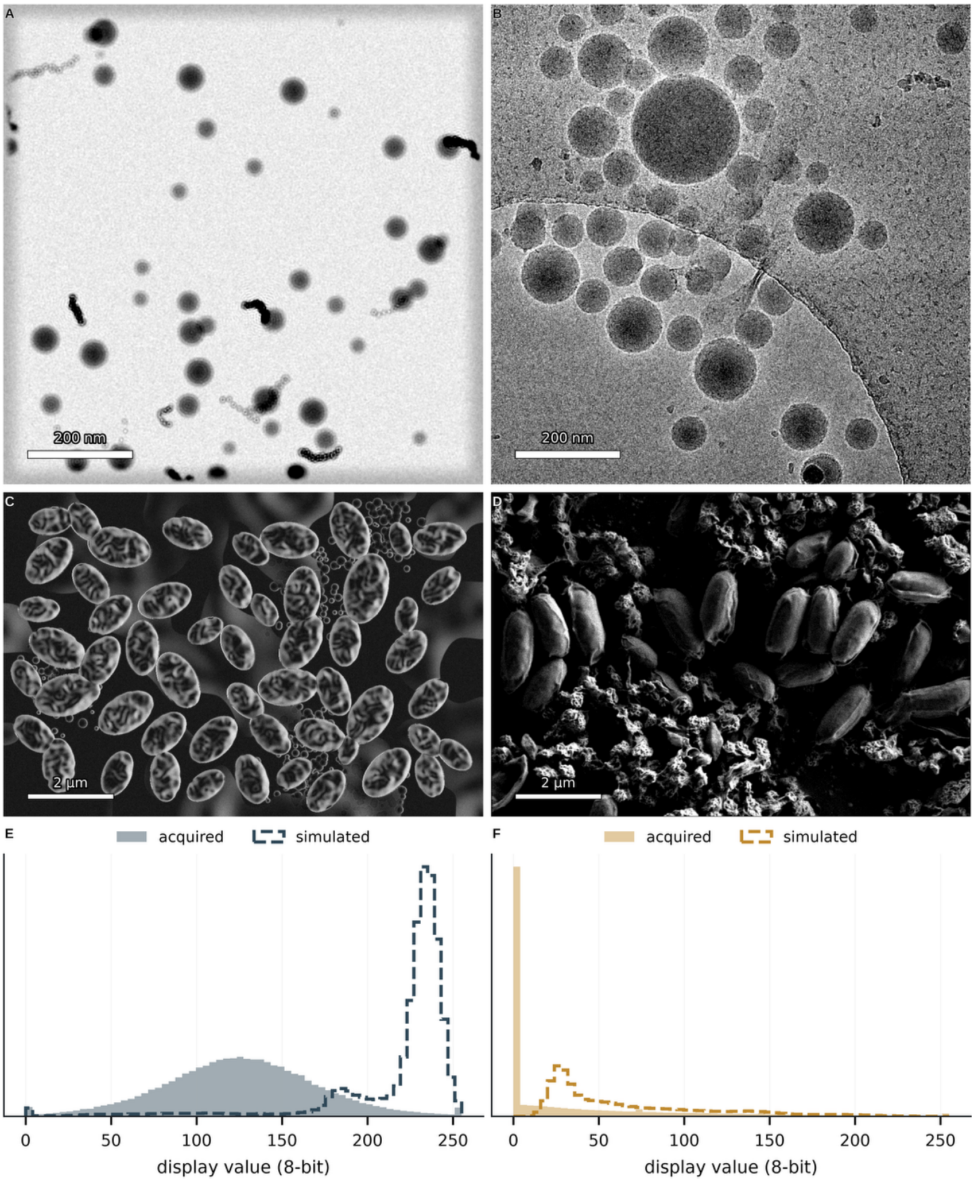
Supplementary Figure S3: **Training behavior of segmentation models fitted to synthetic data.** (A) TEM box and mask mAP_{50} over 100 epochs. (B) SEM box and mask mAP_{50} over 95 epochs. The training stopped with an early-stopping patience of 25 epochs. (C) Box-regression and segmentation-loss components for the TEM and SEM models. Curves are shown for the recorded training execution. Performance values reported in the manuscript and tables were obtained from `best.pt`, the checkpoint with the highest validation fitness on the model-selection partition.

Supplementary Table S7: Transfer from synthetic training data to acquired TEM and SEM images. Synthetic-domain values are repeated from the held-out test results in Supplementary Table S6 as the within-domain baseline. The "Acquired images" column reports the number used for fine-tuning. The SEM source dataset contained 1,085 total annotations. Model performance was evaluated on the 648 annotations in validation micrographs.

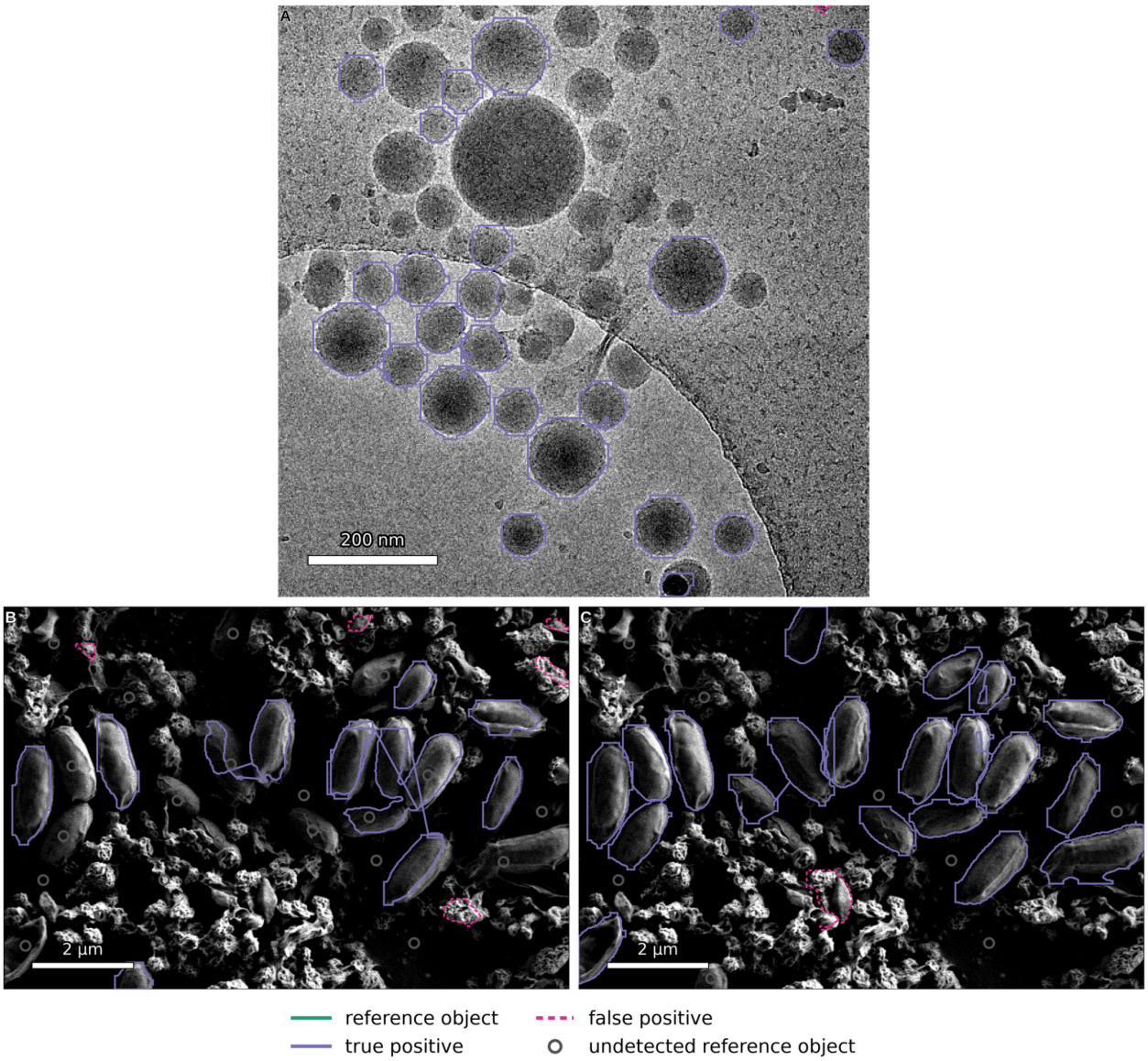
Dataset	Evaluation domain	Acquired images	Recall	Precision
Transmission	synthetic test	0	0.932	0.944
Transmission	Acquired PLGA, 295 verified particles	0	0.329	0.970
Scanning	synthetic test	0	0.988	0.972
Scanning	Acquired SEM reference set	0	0.253	0.432
Scanning	Acquired SEM reference set	4	0.512	0.726



Supplementary Figure S4: **Predictions on synthetic model-selection images.** (A,B) TEM and (C,D) SEM micrographs from the model-selection partition. Two micrographs per modality were selected by proximity to the model-selection-partition median per-micrograph F1 score. Visual appearance was not used for selection. Predictions were scored at a confidence threshold of 0.50. Reference boundaries, true positives, false positives, and undetected reference objects are identified in the unlettered key. The near-ceiling TEM micrograph performance (median F1 $= 0.985$) is a matched-distribution control.



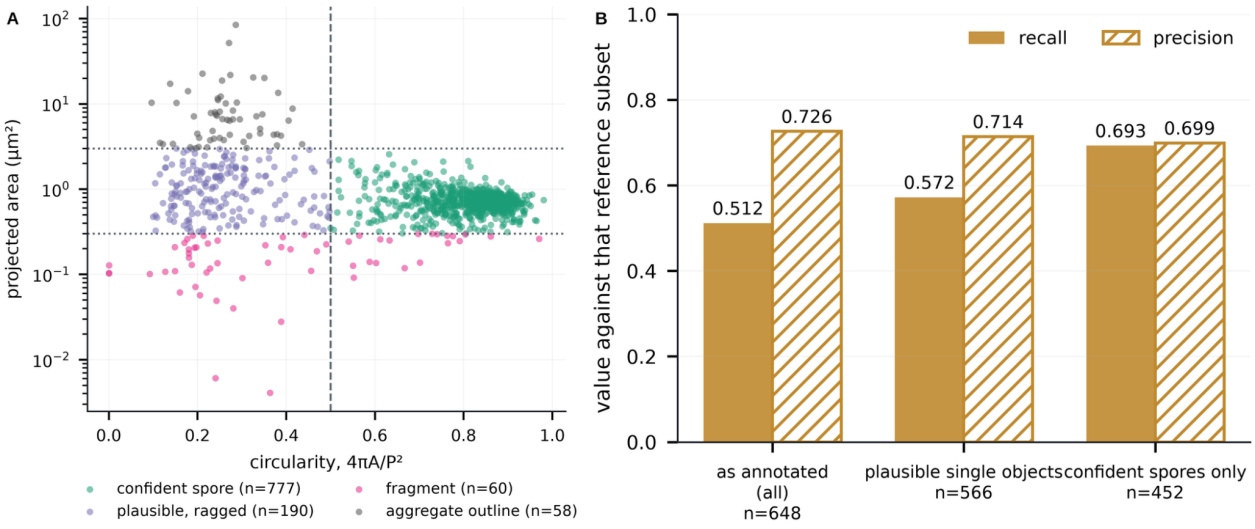
Supplementary Figure S5: **Comparison of simulated and acquired image appearance.** (A,B) TEM and (C,D) SEM examples, with simulated images shown in A and C and acquired images shown in B and D. Images within each modality use the same display window, calculated from the acquired image, and are shown without geometric stretching. (E,F) Pooled 8-bit intensity distributions for the TEM and SEM comparisons, respectively. The filled distributions denote acquired data and dashed curves denote simulated data. For TEM, the acquired and simulated means were 126 and 219 digital numbers (DN), respectively, and 2.1% and 0.8% of pixels were below DN 32.



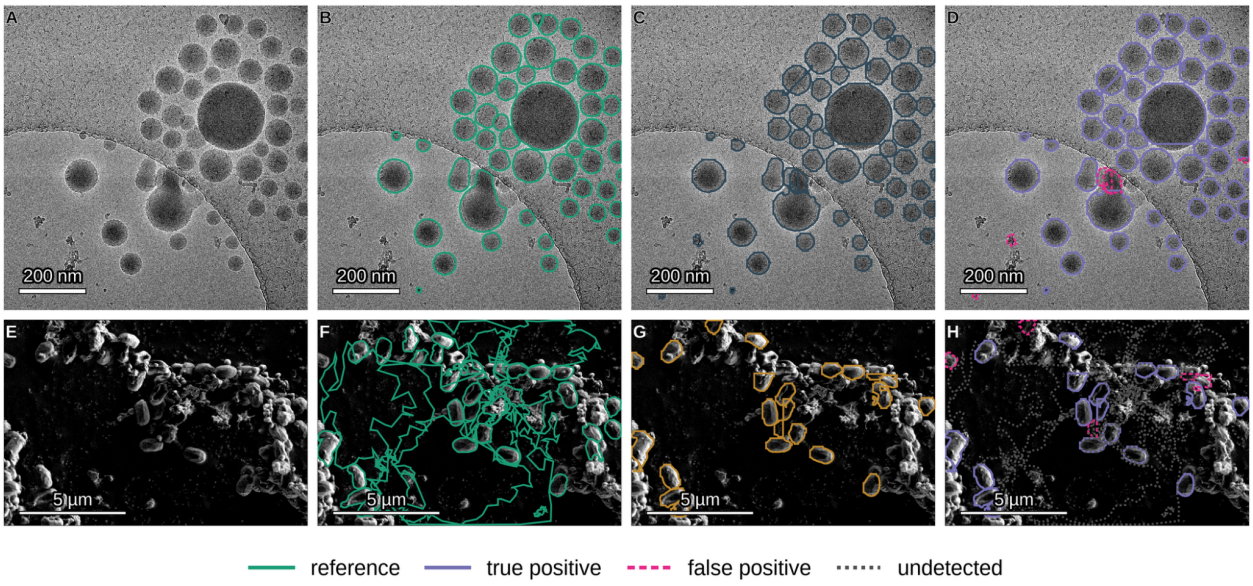
Supplementary Figure S6: **Transfer of synthetic-trained models to acquired images.** Micrographs were selected by proximity to the dataset-median recall. (A) Acquired TEM micrograph evaluated using the simulation-only model: 23 of 60 reference objects were detected, with one false positive and 37 undetected objects (recall = 0.383). No TEM fine-tuning series was performed. (B) Acquired SEM micrograph `4AC2_008` evaluated using the simulation-only model. (C) The same SEM micrograph evaluated after fine-tuning on four acquired micrographs. From B to C, true positives increased from 10 to 20, undetected objects decreased from 25 to 15, and false positives decreased from 5 to 1. The unlettered key defines the reference, true-positive, false-positive, and undetected-object outlines. In the tested SEM micrograph, fine-tuning increased true positives from 10 to 20, reduced undetected objects from 25 to 15, and reduced false positives from 5 to 1.

Supplementary Table S8: Composition of the SEM reference annotations. Categories were assigned using circularity 0.50 and projected-area thresholds of 0.3–3.0 μm^2. The validation micrographs contain 648 of 1,085 annotations and were excluded from fine-tuning. Percentages in the final column are the fractions of each full-set category retained in the validation micrographs.

Reference subset	**Full set,** n	**Held-out validation,** n	**Held-out validation share**
Confident spore	777	452	58%
Plausible, ragged	190	114	60%
Fragment	60	50	83%
Aggregate outline	58	32	55%
Total	1085	648	60%



Supplementary Figure S7: **Reference-set composition changes the apparent model performance.** (A) Projected area versus circularity for all 1,085 SEM reference regions. Dashed thresholds at circularity 0.50 and area 0.3–3.0 μm^2 separate 777 confident spores, 190 plausible ragged spores, 60 fragments, and 58 aggregate outlines. Points represent individual annotated regions. (B) Recall and precision for the same model and predictions scored against successively restricted reference subsets in the various validation micrographs: all annotations ($n = 648$), plausible single objects ($n = 566$), and confident spores only ($n = 452$). Recall increased from 0.512 to 0.693, whereas precision changed from 0.726 to 0.699. The model and predictions were unchanged.



Supplementary Figure S8: **Predictions on validation micrographs after training from reviewed experimental annotations.** (A–D) Cryo-EM input image, reviewed reference annotations, model predictions, and scored overlay, respectively. (E–H) Corresponding SEM panels. The unlettered key identifies reference boundaries, true positives, false positives, and undetected reference objects. Source micrographs were partitioned before tiling. Reference annotations, containment-merging rules, scoring thresholds, and model settings were finalized before evaluation. The validation micrographs were not used for training, model selection, prompt development, or post hoc refinement of the evaluation procedure.

Supplementary Table S10: Experimental TEM and SEM datasets, source-image partitions, training settings, and validation performance. Source micrographs were partitioned before tiling. Checkpoints were selected using the model-selection partition. Precision, recall, and F1 are means ± standard deviation (SD) across three seeds, and the reported metrics were calculated on the validation partition. The independent TEM recall was calculated against a separately generated 235-particle CryoBlob list.

A. Dataset composition and validation performance

Modality	Source images	Training images/tiles	Model selection images/tiles	Validation images/tiles	Annotated objects	Precision	Recall	F1	Box $\mathbf{mAP_{50}}$	Mask $\mathbf{mAP_{50}}$	Mask $\mathbf{mAP_{50–95}}$
Cryo-TEM	30	19/155	3/26	8/71	881	0.893 ± 0.027	0.920 ± 0.009	0.906 ± 0.013	0.963	0.957	0.775
SEM	26	14/84	4/24	8/48	1,085	0.750 ± 0.001	0.534 ± 0.023	0.624 ± 0.016	0.629	0.580	0.355

B. Training and execution settings

Setting	Cryo-TEM	SEM
Architecture	YOLO11s-seg, natural-image pretrained	YOLO11s-seg, natural-image pretrained
Training	100 epochs; batch 16; 512-pixel input; seeds 0, 1, and 2	100 epochs; batch 16; 512-pixel input; seeds 0, 1, and 2
Resolved optimizer	AdamW; learning rate 0.002; momentum 0.9	AdamW; learning rate 0.002; momentum 0.9
Augmentation	HSV 0.015/0.7/0.4; translation 0.1; scale 0.5; horizontal flip 0.5; mosaic 1.0, disabled for the final 10 epochs; erasing 0.4; RandAugment; no rotation, shear, perspective, vertical flip, mixup, or copy-paste.	
Training time	4.8 min	3.8 min
Inference time	1,305 ms per micrograph; 21 ms per tile	830 ms per micrograph; 29 ms per tile
Hardware	Tesla V100-SXM3-32GB GPU	

Supplementary Table S11: Summary of 613 automated calibration-retention tests. The listed defects were corrected before the post-fix result was calculated.

Test or defect	**Evaluation**	**Post-fix result**
Pixel-size retention	Loading, metadata restoration, binning, measurement, and export	maximum relative error 0
OME-TIFF metadata	Pixel calibration was not read	corrected
Sidecar plus binning	Restored pixel size was not rescaled after binning	corrected
Measurement export cache	Export could revert to $1.0\ \mathrm{nm\,px^{-1}}$ when the image was absent from the cache	corrected
COCO polygon dimensions	Source-mask dimensions compared with simplified polygon dimensions	maximum difference 2.89%

Supplementary Table S12: Audit of annotation provenance and downstream records across 26 SEM source micrographs. Percentages for accepted or corrected annotations use the 1,375 final annotation records as the denominator. Rejected candidates remain in the review log but are not part of the final annotation set (percentages are not reported for those rows). The measurement export contains one row per final annotation. Final measurements and statistical outputs are downstream records and are listed separately from annotation-origin counts.

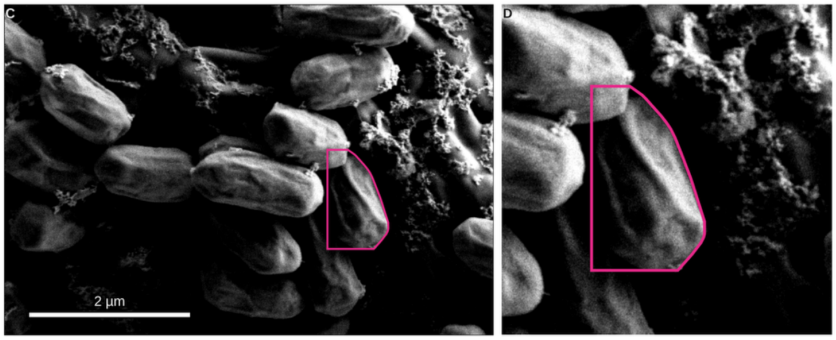
Record category	**Count**	**Percentage**	**Linked record**
Manual annotations	0	0.0%	Source image
Prompt-generated, accepted	48	3.5%	Prompt and model
Prompt-generated, corrected	3	0.2%	Parent annotation
Prompt-generated, rejected	262	–	Review log
Trained-model, accepted	178	12.9%	Checkpoint
Trained-model, corrected	61	4.4%	Checkpoint and parent annotation
Trained-model, rejected	1,818	–	Review log
Imported	1,085	78.9%	Import record
Final measurements	1,375	–	Annotation and calibration
Statistical outputs	1	–	Measurement set

A

plotted value	3A28 ECD 971.5 nm
statistical analysis	A-f012dff3164f Kruskal-Wallis p = 6.75e-04 n = 191
measurement record	M-c09b9e0984e5 area 741233 nm2 circularity 0.91
annotation	b78baf8cbc004387a3b216861b14da81
origin	imported_sidecar batch unedited
geometry hash	created 32bd87ec1cec0d7e current 32bd87ec1cec0d7e
model / prompt	none: not model-generated
dataset version	05e34fdc50931cd5 partitions train 14, val 4, test 8
source micrograph	3A28_004.png
source sha256	6d8fdd3bafd1125bf4f96ed0bd9d803a18d75040f8893ceb550533e8bfeddb26
calibration	9.452 nm/px Zeiss CZ_SEM ap_image_pixel_size (vendor TIFF tag)

B

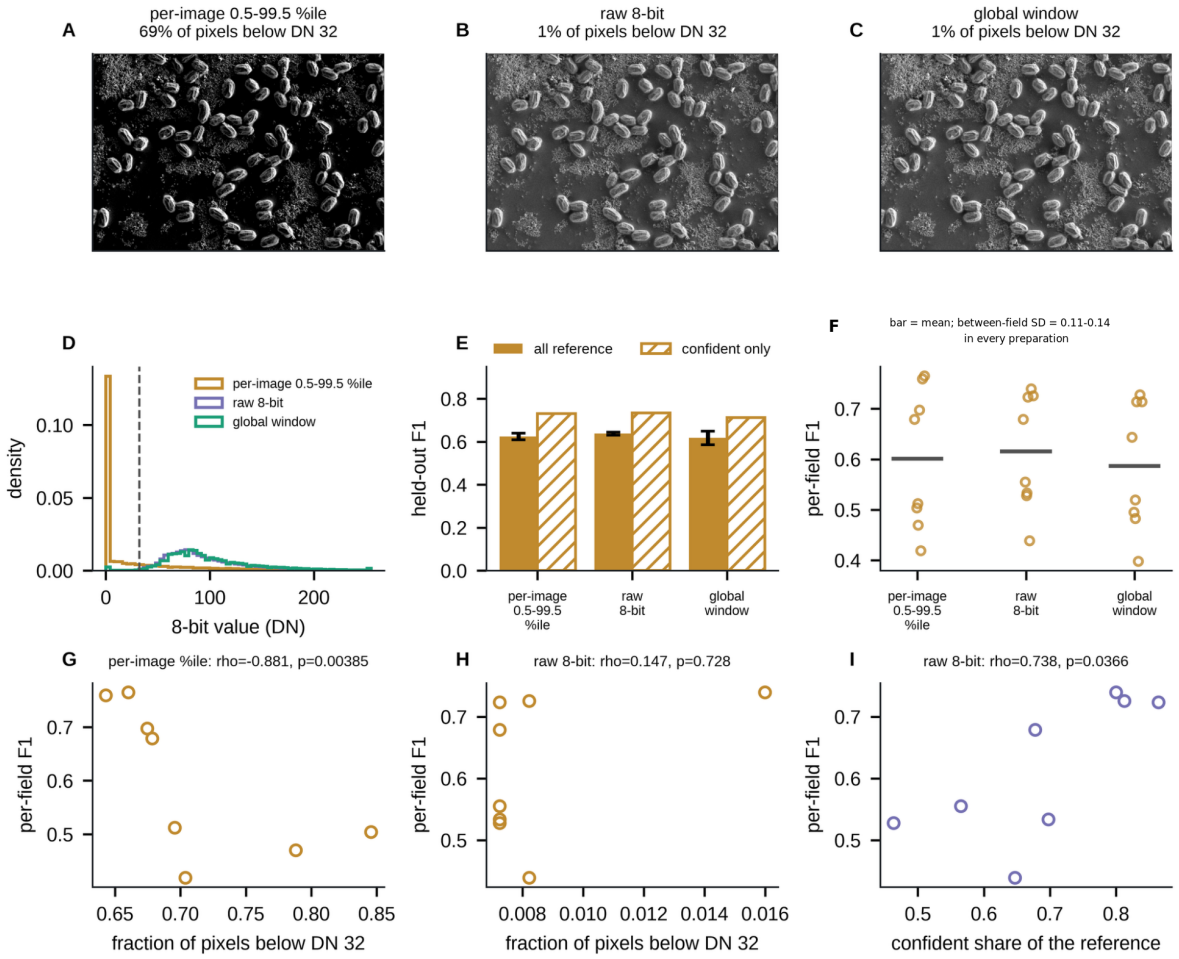
plotted value	3A28 ECD 971.4 nm
statistical analysis	A-f012dff3164f Kruskal-Wallis p = 6.75e-04 n = 191
measurement record	M-e75a97f25fe6 area 741161 nm2 circularity 0.76
annotation	e34564c37ea74d00b7900b639ca8fc3f
origin	yolo direct_gui edited after creation
geometry hash	created 3d049fdf1b7315c8 current a50a7bde668fa9a6
checkpoint sha256	ac4f46852de0c2c41fe670e01212fdc5cc0dd17adab5b3f7d3accb5212f2eb45
training run	seed 0 yolo11s-seg.pt 100 epochs batch 16 512 px
dataset version	05e34fdc50931cd5 partitions train 14, val 4, test 8
source micrograph	3A28_012.png
source sha256	477cb1766a1158c8d155a12bdac8dc79b693683e42ae12c730154e26978adc75
calibration	5.920 nm/px Zeiss CZ_SEM ap_image_pixel_size (vendor TIFF tag)



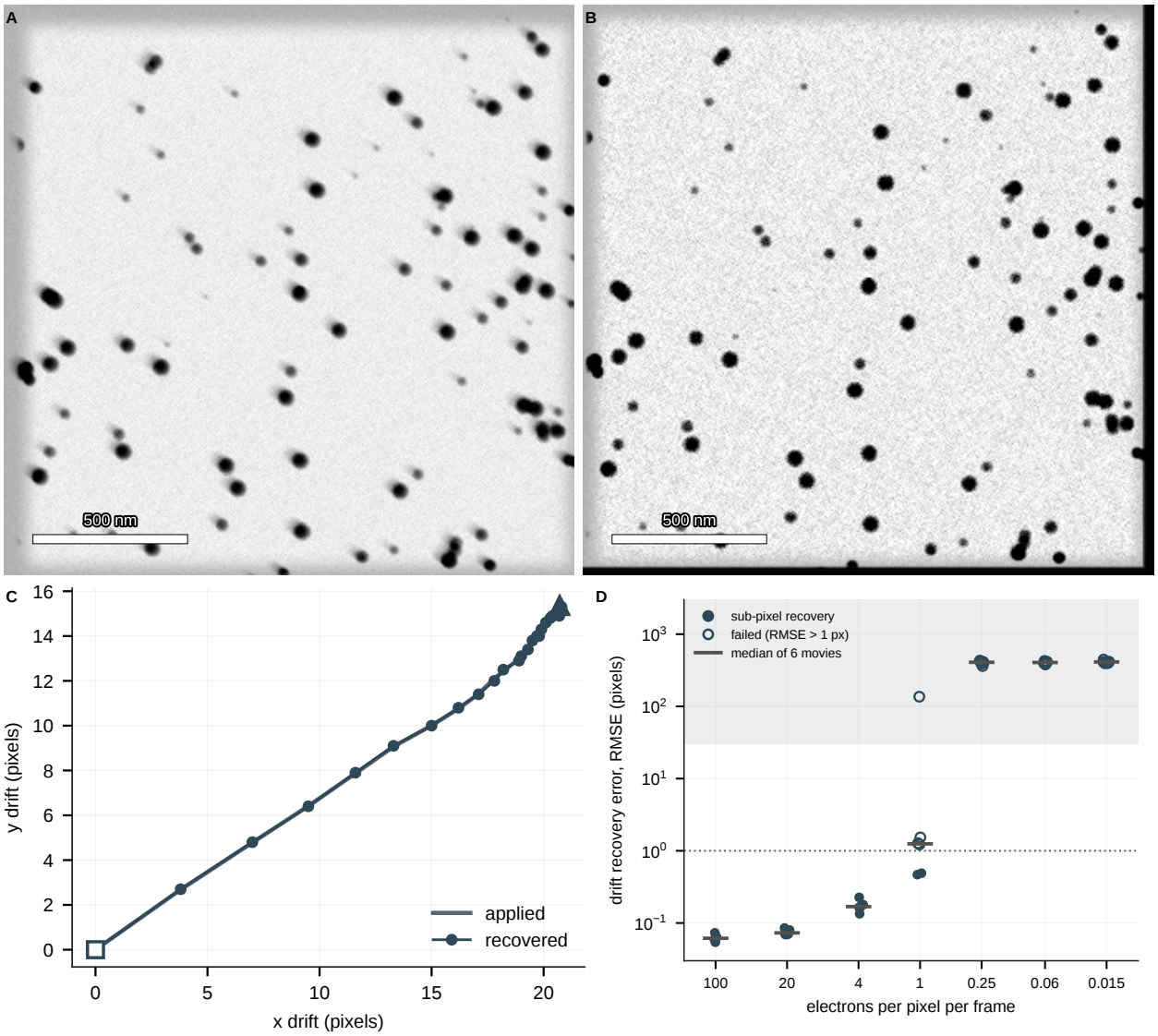
Supplementary Figure S9: **Examples of measurement provenance retained by ACORN.** (A) Provenance chain for the object nearest its group median, linking the reported measurement to the measurement record, annotation, source micrograph, software revision, and spatial calibration. (B) Corresponding chain for the nearest-median object produced by a trained model, including the checkpoint identifier and training-run record. (C,D) Source-image regions and stored mask geometries for the records shown in A and B. Values, identifiers, and hashes are displayed as stored. Each reported number is linked to its mask, source image, calibration, and code version.

Supplementary Table S13: Effect of model initialization and the number of annotated acquired SEM micrographs on mask mAP_{50} across 18 validation micrographs in the legacy `real_ft` experiment. This partition is separate from the later 26-micrograph SEM benchmark. At two and four micrographs, each row summarizes four independently resampled micrograph sets. At eight micrographs, each row summarizes 10 executions using the same fixed eight-image pool. The simulation-initialized executions at eight micrographs produced identical predictions and $\mathrm{mAP}_{50} = 0.529393$ under the deterministic training configuration. All simulation-initialized conditions reused one simulation-pretrained checkpoint. These runs do not constitute independent simulation-pretraining replicates. SD denotes standard deviation.

Initialization	**Acquired micrographs**	**Runs**	**Replicate type**	**Mean**	**SD**	**Range**	**Median**
Simulation	2	4	Resampled sets	0.2912	0.1760	0.1343–0.4561	0.2872
Natural image	2	4	Resampled sets	0.4612	0.0452	0.3981–0.5041	0.4714
Simulation	4	4	Resampled sets	0.5103	0.0128	0.4945–0.5259	0.5105
Natural image	4	4	Resampled sets	0.5300	0.0131	0.5112–0.5391	0.5349
Simulation	8	10	Fixed image pool	0.5294	0 (exact)	0.5294–0.5294	0.5294
Natural image	8	10	Fixed image pool	0.4810	0.0364	0.4292–0.5411	0.4759



Supplementary Figure S10: **Effects of SEM intensity preparation and micrograph composition.** (A–C) The same SEM micrograph after per-image 0.5–99.5 percentile stretching, retention of the acquired 8-bit values, and application of one global window, respectively. (D) Corresponding pooled intensity distributions. (E) F1 scored against all references and against confident spores only. Error bars show the standard deviation (SD) across three training seeds. (F) Per-micrograph F1 for eight validation micrographs under each preparation. The points show micrographs and horizontal bars show means. (G,H) Per-micrograph F1 versus the fraction of pixels below DN 32 for the per-image and raw preparations. (I) Per-micrograph F1 under raw preparation versus the confident-spore fraction of the reference set. Between-micrograph SD was 0.14, 0.11, and 0.13 for the per-image, raw, and global-window preparations, respectively. Spearman associations were $\rho = -0.881$ ($p = 0.0039$), $\rho = 0.147$ ($p = 0.73$), and $\rho = 0.098$ ($p = 0.82$) for dark-pixel fraction under the per-image, raw, and global-window preparations. The association between raw-preparation F1 and confident-spore fraction was $\rho = 0.738$ ($p = 0.0366$).



Supplementary Figure S11: **Motion-correction performance benchmark.** (A) Frame average without motion correction. (B) The same movie after motion correction. Panels A and B use a single 0.5–99.5 percentile window taken from the uncorrected average. (C) Recovered drift trajectory compared with the known applied trajectory. (D) Drift-recovery error as a function of per-frame dose. All six movies per condition are shown. Filled markers indicate subpixel recovery and open markers indicate failures (defined as RMSE greater than one pixel) and horizontal bars are condition medians.

Supplementary Table S14: Summary of the replicated motion-correction evaluation. Forty-eight independent movies were generated, six per dose condition and 24 frames per movie. Dose in $\mathrm{e^{-}\,\AA^{-2}}$ was calculated using an 18 Å/pixel calibration. Alignment succeeded when the trajectory root-mean- square error (RMSE) was less than 1 pixel. Sharpness was defined as the variance of the vertical finite-difference image gradient. Sharpness gain was calculated as $100[S(I_{\mathrm{corrected}})/S(I_{\mathrm{uncorrected}}) - 1]$ percent. RMSE and sharpness changes are reported as mean $\pm$ standard deviation (SD).

Condition	**Dose e$^-$/px/frame**	**Dose e$^-$/Å^2/frame**	**Total dose e$^-$/Å^2**	n	**RMSE mean $\pm$ SD (px)**	**RMSE median (px)**	**RMSE range (px)**	**Sharpness change (%)**	**Success**	**Failure**	**Success fraction**
Nominal	300	0.926	22.2	6	0.064 ± 0.004	0.063	0.057–0.069	309 ± 51	6	0	1.000
100	100	0.309	7.41	6	0.062 ± 0.007	0.061	0.054–0.073	238 ± 32	6	0	1.000
20	20	0.0617	1.48	6	0.075 ± 0.007	0.073	0.069–0.085	95 ± 11	6	0	1.000
4	4	0.0123	0.296	6	0.174 ± 0.030	0.169	0.134–0.226	10 ± 4	6	0	1.000
1	1	0.00309	0.0741	6	23.527 ± 55.190	1.248	0.467–136.180	-14 ± 2	2	4	0.333
0.25	0.25	0.000772	0.0185	6	402.323 ± 23.965	408.152	362.297–429.779	-43 ± 2	0	6	0.000
0.06	0.06	0.000185	0.00444	6	402.764 ± 17.558	405.812	382.808–425.475	-43 ± 2	0	6	0.000
0.015	0.015	0.0000463	0.00111	6	412.692 ± 17.641	412.209	394.542–444.407	-44 ± 2	0	6	0.000

Supplementary Table S15: SEM validation after changing intensity preparation and source-image partitioning. Aggregate F1 values are mean ± standard deviation (SD) across three training seeds. Between-micrograph SD was calculated across eight validation micrographs after assigning F1 = 0 to a micrograph containing reference objects but no correct detections. Spearman correlations are micrograph-level and exploratory. NA = not applicable.

Preparation	**Pixels below DN 32 (%)**	**Validation F1 mean ± SD**	**Confident-reference F1**	**Training seeds**	**Between-micrograph SD**	**ρ: dark fraction**	**p: dark fraction**	**ρ: confident share**	**p: confident share**
Per-image 0.5–99.5 percentile	68.05	0.6239 ± 0.0158	0.7299	3	0.1389	-0.881	0.00385	0.643	0.0856
Raw acquired 8-bit values	1.87	0.6376 ± 0.0067	0.7338	3	0.1149	0.147	0.728	0.738	0.0366
One global intensity window	1.80	0.6176 ± 0.0314	0.7119	3	0.1282	0.098	0.817	0.643	0.0856
Isolate-stratified split	NA	0.6516 ± 0.0249	0.7292	3	0.2402	NA	NA	–	–

Supplementary Table S16: Physical-measurement error among matched detections on the held-out synthetic test partitions. Predicted and ground-truth instances were assigned one-to-one in descending mask intersection-over-union, requiring matching class and intersection-over-union ≥ 0.5. Matching covered 2,097 of 2,250 TEM particles (93.2%) and 1,069 of 1,083 SEM spores (98.7%). Reported differences reflect segmentation-boundary effects among matched detections. MARE is the mean absolute relative error.

Modality	**Measurement**	**Bias**	**MAE**	**MARE**
TEM	Area (nm^2)	$+189.8$	246.3	21.5%
TEM	Equivalent circular diameter (nm)	$+2.90$	3.60	10.1%
TEM	Feret diameter (nm)	$+12.14$	12.17	33.1%
SEM	Area (nm^2)	$+36{,}024$	58,446	7.6%
SEM	Equivalent circular diameter (nm)	$+23.21$	31.48	3.7%
SEM	Feret diameter (nm)	$+65.84$	83.95	7.4%

Supporting Information References

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

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

[3] napari contributors. napari: A multi-dimensional image viewer for python, 2019.

[4] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9:676–682, 2012.

[5] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9:671–675, 2012.

[6] 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.

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

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

[9] Sjors H. W. Scheres. RELION: Implementation of a bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012.

[10] Jasenko Zivanov, Takanori Nakane, Bjorn O. Forsberg, Dari Kimanius, Wim J. H. Hagen, Erik Lindahl, and Sjors H. W. Scheres. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, 7:e42166, 2018.

[11] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14:290–296, 2017.

[12] 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.

[13] 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.

[14] Wenhao Yuan, Bingqing Yao, Shengdong Tan, et al. Generative learning of morphological and contrast heterogeneities for self-supervised electron micrograph segmentation. *npj Computational Materials*, 11:322, 2025.

[15] 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.