LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-842942

Enhancing Nanoparticle Detection in Interferometric Scattering (iSCAT) Microscopy Using Mask R-CNN

M. J. Boyle, Y. E. Goldman, R. J. Composto

December 5, 2022

Journal of Physical Chemistry B

Disclaimer

Enhancing Nanoparticle Detection in Interferometric Scattering (iSCAT) Microscopy Using Mask R-CNN

Michael J. Boyle[1,2] (boyle17@llnl.gov), Yale E. Goldman[3] (goldmany@upenn.edu) and Russell J. Composto[1] * (composto@seas.upenn.edu)

[1]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA, 19104, United States.

[2]Center for Engineered Materials and Manufacturing, Lawrence Livermore National Laboratory, Livermore, CA, 94550, United States

[3]Department of Physiology and Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, PA, 19104, United States.

KEYWORDS: Interferometric Scattering Microscopy, Object Detection, Supervised Machine Learning, Image Processing

ABSTRACT

Interferometric scattering microscopy (iSCAT) is a label-free optical microscopy technique that enables imaging of individual nano-objects such as nanoparticles, viruses, and proteins. Essential to this technique is the suppression of background scattering and identification of signals from nano-objects. In the presence of substrates with high roughness, scattering heterogeneities in the background, when coupled with tiny stage movements, cause features in the background to manifest in background-suppressed iSCAT images. Traditional computer vision algorithms detect these background features as particles, limiting the accuracy of object detection in iSCAT experiments. Here, we present a pathway to improve particle detection in such situations using supervised machine learning via a mask region based convolutional neural network (mask R-CNN). Using a model iSCAT experiment of 19.2 nm gold nanoparticles adsorbing to a rough layer-by-layer polyelectrolyte film, we develop a method to generate labeled datasets using experimental background images and simulated particle signals, and train mask R-CNN using limited computational resources via transfer learning. We then compare the performance of mask R-CNN trained with and without inclusion of experimental backgrounds in the dataset against a traditional computer vision object detection algorithm, Haar-like feature detection, by analyzing data from the model experiment. Results demonstrate that including representative backgrounds in training datasets improved mask R-CNN in differentiating between background and particle signals, and elevated performance by markedly reducing false positives. The methodology for creating a labeled dataset with representative experimental backgrounds and simulated signals facilitates the application of machine learning in iSCAT experiments with strong background scattering, and thus provides a useful workflow for future researchers to improve their image processing capabilities.

INTRODUCTION

In the past two decades,[1] interferometric scattering (iSCAT) microscopy has enabled imaging and tracking of nano-objects such as gold nanoparticles,[2–5] viruses,[6] and individual proteins[7–9] for dynamic studies on cell membranes,[5] molecular motors,[10] and quantitative mass/size measurements.[11–13] As a scattering-based technique, imaging of nano-objects in iSCAT experiments does not require fluorescent labeling.[14] This advantage mitigates challenges inherent to fluorescence-based techniques such as photobleaching and photoblinking.[14] Fluorescent labels on the surfaces of nano-objects also alter their surface chemistry, rendering iSCAT advantageous in studies where surface chemistry dominates (e.g., adsorption experiments, interactions within a cell, diffusion along surfaces). Given these advantages, the contributions of iSCAT to the biological and physical science community are continually growing.

The label-free capabilities of iSCAT also represent its greatest limitation.[15] All materials with a different refractive index than the surrounding medium will scatter light and convolute signals of interest with the background. Consequently, extracting meaningful information from iSCAT images is a two-step process: (1) identifying and suppressing the background and (2) detecting objects of interest in background-suppressed images. While numerous background extraction and suppression techniques have been developed though optical enhancements[12,16,17] and software algorithms,[8,12,14,18] stage movements, on the order of tens nanometers (comparable to the pixel size in many iSCAT configurations), are difficult to eliminate and cause background features to appear even in background-suppressed images.[15] This problem is especially evident when the background contains regions of high roughness or refractive index heterogeneity, and limits the ability of iSCAT to detect scatterers in complex environments. Here, background features that produce signals due to stage drift or vibrations appear as objects with similar intensity and

morphology as objects of interest in background-suppressed images. Experimenters are typically required to perform time-consuming, intricate filtering of detections with similar characteristics. The method could therefore benefit from improvements in step two of the analysis process through improved computer vison strategies to detect particles in complex iSCAT images where less effort has been focused to date.

Convolutional neural networks (CNN) in the field of supervised machine learning computer vision have emerged as a powerful means to improve performance in the problem of object detection in scientific images relative to traditional edge and thresholding computer vision algorithms.[19,20] Still, employing supervised machine learning in object detection currently requires a labeled dataset (i.e., known truths of object locations) that is representative of experimental images to use in training a CNN.[21,22] In a recent publication, Newby[23] *et. al* circumvented this limitation by generating a dataset consisting of synthetic fluorescent microscopy images by mathematically modeling both background and particle signals. However, in background-suppressed iSCAT images, the background is comparatively more complex and difficult to simulate. This makes creating simulated datasets to use in training a CNN for object identification in iSCAT images challenging. Compounding the problem, training CNNs from randomly initialized weights requires large datasets and great amounts of computing power.[20] To facilitate the use of machine learning in iSCAT object detection, both these obstacles are addressed in this work.

We demonstrate a machine learning workflow to improve object detection in complex background-suppressed iSCAT images. To accomplish this, we establish a procedure for creating labeled datasets containing synthetic particle signals with and without real experimental background-suppressed images from a model iSCAT experiment of 19.2 nm gold nanoparticles

adsorbing to a rough layer-by-layer polyelectrolyte film. We then train a mask region-based convolutional neural network (mask R-CNN)[20,23] in under one hour using transfer learning[24] with readily accessible computational resources using the labeled datasets. By analyzing the model iSCAT experiment, we test the performance of mask R-CNN (trained with and without experimental backgrounds in the dataset) in object detection and classification against a Haar-like feature image segmentation algorithm, an edge detection-thresholding algorithm previously used to detect objects in iSCAT images.[11,25] Results highlight the improved performance in object detection via a reduction of false positive detections **[**precision improved from 80.2% (Haar) to 96.5% (mask-RCNN)] and the importance of including representative experimental backgrounds in datasets. The result is not an automated, catch-all machine learning network to use in iSCAT image processing, but rather it shows how to create datasets and optimize a CNN via transfer learning to improve object detection in experimental data. The improved analysis technique should expand capability of iSCAT to detect scattering objects in situations where background scattering is complex.

iSCAT Principles: Stage Movements Hinder Background Suppression

The theory behind iSCAT imaging has been reviewed extensively.[15,26,27] We provide a summary of iSCAT operating principles below to highlight the benefit of enhanced object detection methods in expanding the application of iSCAT to substrates with strong scattering background features due to roughness and/or areas of varying refractive index.

The contrast in an iSCAT experiment is the result of interference between light scattered from nano-objects (e.g., nanoparticles, proteins, viruses) and light from a reference source, commonly light reflected back at the substrate-solution interface from a normally incident coherent light source. The intensity of the signal detected is determined by the superposition of the scattered and reflected light

$$I_{iSCAT} \propto |E_r + E_s|^2 = E_i^2[\, r^2 + s^2 + 2r|s|\cos(\phi)] \qquad (1)$$

where I_{iSCAT} is the intensity at the detector, E_r, E_S, and E_i are the reflected, scattered, and incident electric fields, respectively, r is the fraction of E_i reflected, s is the fraction of E_i scattered, and ϕ is the phase difference between the reflected and scattered light due to a difference in optical path length. According to Mie scattering theory (applicable when the wavelength of light is of the order of the nanoparticle size), the magnitude of s scales linearly with the volume of the particle (d_{NP}^3),[13,28] and in the limit of small scattering objects, reflected light dominates the detected signal making the pure scattering contribution negligible ($r \gg s$).[28] Equation 1 then reduces to

$$I_{iSCAT} = E_i^2[\, r^2 + 2r|s|\cos(\phi)] \qquad (2)$$

Equation 2 illustrates the power of iSCAT and highlights its limitations. In comparison to pure scattering based contrast techniques (i.e. dark field microscopy) where the scattering term for small objects decreases with the volume of the particle squared,[28] the scattering term in an iSCAT signal decreases linearly with the particle volume. This is what lowers the detection threshold for label-free imaging of nanoparticles, viruses, and individual proteins.[9,11] However, because all interfaces with a difference in refractive index scatter light, backgrounds in iSCAT experiments often scatter light at stronger intensities than the nanoparticles or proteins of interest. To generate sufficient contrast, C_{iSCAT}, for analysis, strongly varying background features are typically suppressed by dividing or subtracting a reference background image from experimental images. This background suppression, the mechanisms of which are discussed in the Methods section, is described mathematically as

$$C_{iSCAT} = \frac{I_{iSCAT}}{I_{bkg}} = \frac{E_i^2\left[\, r_{image}^2 + 2r_{image}|s|\cos(\phi)\right]}{E_i^2 r_{bkg}^2} \approx 1 - \frac{2|s|\cos(\phi)}{r} \qquad (3)$$

where r_{bkg} represents the fraction of E_i reflected by the optics and substrate that forms the background reference from an earlier frame capture, and r_{image} represents the fraction of E_i

reflected that forms the image currently being analyzed. If the imaging stage and optical components are stable, $r_{bkg} = r_{image}$, and all background reference features would be suppressed by division resulting in the solution to equation 3 at the right. However, in practice, drift or vibrations on the order of tens of nanometers (i.e., comparable to the pixel dimension) are typical for optical microscopes.[15,29] These stage movements cause r_{bkg} to differ from r_{image}, and scattering from the substrate contributes significant features to background-suppressed images. These substrate features may have morphologies resembling scattering features from nano-objects, and the essential step of detecting objects of interest in background-suppressed iSCAT images becomes more difficult as background complexity increases. The challenges presented by the presence of particle-like background features in background-suppressed iSCAT images motivates our application of machine learning to aid in particle detection in iSCAT experiments with complex backgrounds.

METHODS

Experimental Details

150 μm thick glass coverslips were coated with a layer-by-layer (LbL) film (dry thickness = 60 nm, wet thickness = 85 nm) composed of 11 alternating polycation [poly (allylamine hydrochloride)] and polyanion [poly(acrylic acid)] layers. The LbL film was capped by the polycation leaving a positive zeta potential of +14.7 mV on the substrate when exposed to 1 mM sodium phosphate buffer at a pH of 7.0. (See section S1 for material information and section S2 for substrate preparation and characterization). LbL films were characterized using atomic force microscopy (AFM, figure S3 and S4), and shown to increase the roughness of the substrate by a factor of 4 when compared to uncoated glass coverslips used in other iSCAT experiments ($R_{rms,LBL}$ = 3.2 nm, $R_{rms,glass}$ = 0.8 nm, *Area* = 10 μm x 10 μm).

Using a flow cell (figure S5), a 200 pM solution of monodisperse gold nanoparticles (AuNPs, see section S4 for solution preparation and characterization) with a mean diameter of 19.2 nm in 1 mM sodium phosphate buffer, pH of 7.0, was exposed to the LbL film on an inverted iSCAT microscope (see section S5 for iSCAT instrumentation and imaging details). At this pH, the AuNPs had a negative zeta potential of -58 mV because of their carboxy-terminated ligands and electrostatic interactions drove irreversible adsorption of the AuNPs to the positively charged LbL coated glass coverslip over the course of a 4 hour experiment. At 30 minute intervals, iSCAT images were collected for ~1 minute at a frame rate of 184 fps (~5.44 ms between frames). The pixel size of the iSCAT setup was measured using a stage micrometer to be 50 nm/pixel, giving the 256 x 256 pixel images a field of view of 12.8 μm x 12.8 μm. This resulted in an experimental dataset consisting of 108 videos of 1000 frames each for a total of 108,000 iSCAT images.

Image Flattening

Raw iSCAT videos were processed in two operations: flattening followed by ratiometric processing. The flattening techniques described here were detailed previously[28] and implemented using custom python scripts. In the first flattening step, a static background image was collected by laterally translating the sample stage during data collection in a Lissajous pattern using piezoelectric motors and calculated as the pixelwise median of 100 images collected during that movement. Each raw iSCAT image was divided by its corresponding static background image to remove spurious features inherent to the optical configuration. Next, a median filtered image containing only features larger than the scattering features of the nanoparticles was calculated by convoluting the image from the first flattening step with a kernel of 21 pixels and calculating the median of pixels in that kernel. The resulting image from the first flattening step was then divided by the median filtered image yielding flattened images with a mean background value of 1. All

subsequent ratiometric processing was done on these flattened images. An example of the conversion of a raw iSCAT image to a flattened image is shown in figure S7.

Ratiometric Processing

After flattening, we employed ratiometric processing[12] to suppress the background. By balancing images that follow a nanoparticle adsorption event with those that precede the event, ratiometric processing capitalizes on time-dependent variations in scattering signals during the collection of iSCAT videos to isolate adsorption events in background suppressed images. In doing so, it enables the measurement of the scattering contrast and the binding times of analytes, and has been previously used to detect nanoparticles[13] and proteins.[11]

To apply ratiometric processing to detect nanoparticle adsorption events, we began by defining a time-binning window of N_{ratio} = 5 frames. N_{ratio} effectively defined the minimum residence time for a particle to be considered adsorbed as a particle that remained attached to the surface for at least $2N_{ratio}$ frames. In our experiments, N_{ratio} = 5 frames defined the minimum residence to be $2N_{ratio}$ x *frame interval*, or 54.3 ms. Faster frame intervals would have allowed for the analysis of shorter residence times, but in our experiments, we did not see any desorption events, indicating that this minimum residence time was sufficient.

We calculated ratiometric images as follows. For a given frame *i,* two batches of sequential images were averaged to produce two images $I_1(i)$ and $I_2(i)$, where $I_1(i)$ corresponded to the pixelwise average of images i to N_{ratio} and $I_2(i)$ to images $i + 1 + N_{ratio}$ to $2N_{ratio}$. After normalizing $I_2(i)$ and $I_1(i)$ by dividing each average image by its mean, we then divide $\hat{I}_2(i)/\hat{I}_1(i)$ to obtain ratiometric images, $I_{\text{ratio}}(i)$. Frames were then incremented across each video one frame at a time, creating a new movie consisting of ratiometric images. As frames are incremented forward in time, the scattering contrast of adsorbing particles increases in magnitude reaching a maximum when

the adsorption event is located between the two frame batches, and decreases back to the background value of 1 thereafter. Adsorption events in in our configuration destructively interfere with the reflected light and manifest in dark spots on a grey background. The reverse process, desorption, if present, would present as bright spots on a grey background. In this way, by finding the point of maximum scattering contrast magnitude, the scattering contrast and occurrence time for adsorption and desorption events can be precisely quantified. An illustration of ratiometric processing is illustrated in figure S8 (a)-(c).

Particle Detection Using the Haar Method

Following ratiometric processing, particle signals or point spread functions (PSFs) were detected using traditional and machine learning object detection algorithms. The Haar-like feature algorithm was implemented in python following details described in ref. 11 and 30. First, ratiometric images were convoluted with a 2D-Gaussian in equation 4

$$I_{PSF}(x,y) = A\exp\left(-\left(\frac{(x-x_0)^2}{2\sigma} + \frac{(y-y_0)^2}{2\sigma}\right)\right) + B \qquad (4)$$

where A is the contrast amplitude, x_0 and y_0 are the center position, B is the background, and σ is the standard deviation. The parameters of the 2D-Gaussian used for the convolution were determined empirically by fitting a particle PSF in a ratiometric image. Next, Haar feature scores were calculated for each pixel with higher scores corresponding to vertical edges, horizontal edges, and circular features using three kernels each 9 x 9 pixels in size. The pixelwise Haar scores were averaged between these three features and weakly thresholded to remove pixels with Haar feature scores lower than 0.20. From the weakly thresholded images, only thresholded pixels with 4 neighbors were kept as candidate pixels that belong to a particle signal. The probability a pixel belongs to particle was then calculated as the fraction of pixels that satisfied all threshold and neighbor candidacy requirements within a 7 x 7 kernel of neighboring pixels. If this probability

was greater than 0.3, the pixel was classified as belonging to a particle, resulting in a segmented binary image that was then labeled in object detection. Although the Haar-like feature algorithm was used as a control in this study, it would be valuable in future studies to test other state-of-the-art object detection methods, such as the circular Hough transform used by Melo[13] *et. al.,* to see how performance compares to Mask R-CNN when the background scattering is high.

Particle Detection using Mask R-CNN

Mask R-CNN, described in detail by He[20] et al., detects objects in images in four main stages. First, features are detected and mapped to pixels using a CNN. From these feature maps, bounding boxes around regions of interest (ROIs) are proposed. Then, simultaneously, bounding boxes around ROIs are refined, regions are classified, and instance-level segmentation masks within the refined bounding boxes are generated as the final outputs. In our implementation, we capitalized on the inference output of bounding boxes and classifications to perform Gaussian fitting and particle tracking in subsequent steps.

To implement mask R-CNN for object detection in iSCAT images, Matterport's implementation[23] of mask R-CNN in TensorFlow and Keras was used with minor hyperparameter modifications. Namely, a ResNet50 backbone was used, and image dimensions were constrained to the size of our iSCAT images, 256 x 256 pixels. For training, all ROIs with a detection confidence below 0.7 were rejected. In inference, all ROIs with a detection confidence below 0.9 were rejected. We normalized all input images by rescaling their intensity to a minimum of 0 and maximum of 1, and converted the normalized greyscale images to 8-bit RGB images. While not an exhaustive optimization, only these minor changes were required to obtain good performance in both training and inference. Additionally, maintaining most hyperparameters at their default

values made transfer learning from ImageNet weights easier, reducing the amount of data required for training and shortening the training time on limited computational resources.

Creating Datasets to Train Mask R-CNN

An essential contribution of this work is the methodology for creating labeled image datasets with realistic experimental backgrounds to use in training a neural network for object detection. A custom python script was used to generate 500 synthetic dataset images. First, a particle image was generated using a PSF approximation in the form of a 2D-Gaussian (equation 4) to generate synthetic nanoparticle signals approximating those measured in iSCAT on an image with a background of 0. In experiments with less background roughness, secondary interference fringes beyond the central lobe maybe visible, and Bessel functions, which more accurately approximate the interference of a plane wave with a spherical wave, may serve as better PSF models for simulated particle signals. PSF parameters (listed in table 1) were empirically determined by fitting 10 particles from ratiometric images. Notably, an amplitude (A) at the high end of the measured scattering contrast from our experiments was selected to increase the contrast between synthetic particle signals and background features. We found this contrast vital to ensure good performance during inference. Dark particles were assigned a negative amplitude, while bright particles were assigned a positive amplitude in the 2D-Gaussian. Particles were randomly positioned within an image of the same size as the iSCAT ratiometric images with the minimum spacing between other particles and the border limited, and positions and particle class types were logged. Instance-level masks with corresponding class labels were then generated using particle positions and a circular mask of defined size to serve as the ground truth in training.

Table 1. Parameters used to create synthetic particle images for our dataset.

Parameter	Value
A (a.u.)	0.06
σ (pixels)	2.00
Particles per image	10-15
Minimum spacing between particles (pixels)	40
Minimum particle centroid distance to border (pixels)	5
Mask diameter (pixels)	21
Particle classes	Dark, Bright
$\sigma_{shot\ noise}$ (a.u.)	0.001 - 0.004
Training Image Number	350
Test Image Number	150

To add variance to the dataset and make it more realistic, a shot noise image of Gaussian noise with a mean of zero and a given standard deviation, $\sigma_{shot\ noise}$, was created. Next, a real background from the ratiometric experimental images was sampled as follows. One image from every third video (1,4,7,…,106) in the experiment was randomly selected to capture representations of the background over the course of the iSCAT experiment. To ensure background images contain no particles, images were visually inspected and resampled until no particle PSFs were present in any background images resulting in a set of 36 particle-free ratiometric background images. When the dataset was generated, one image from this set was randomly selected to serve as the background for that image. Figure 1 shows how the particle image, shot noise, and randomly sampled background (top row) were summed to create the dataset image (bottom, left) and corresponding labels for each class (bottom, middle and right).

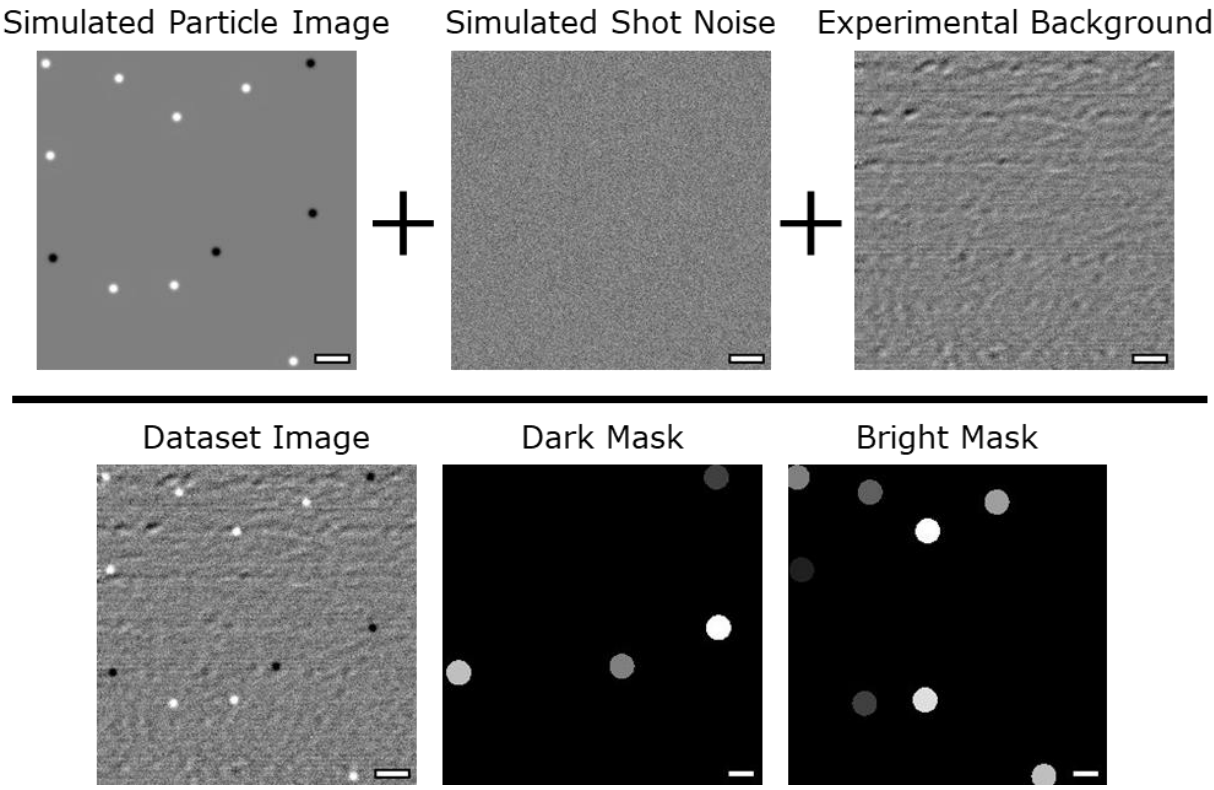


Figure 1. A visual representation of the process of creating labeled datasets for training mask R-CNN via transfer learning. The dataset image (bottom, left) was created by summing the simulated particle image (containing randomly positioned synthetic PSFs), simulated shot noise image, and a randomly sampled experimental ratiometric background image. Using stored particle positions from the simulated particle image, instance-level masks (bottom, middle and right) for dark and bright particle classes served as labels for training and testing. All scale bars are 1 μm but were not included in the dataset images.

Training Mask R-CNN

Training via transfer learning was implemented using configurations detailed by Abdulla[23] with minimal modifications. In brief, network weights were initialized from a mask R-CNN network trained on the ImageNet dataset. Three new classes ("background", "bright", and "dark") were defined for the generated dataset images and corresponding masks. Images in the dataset were augmented during training through vertical and horizonal flips, Gaussian blurring, and

scaling to limit overfitting and improve generalization. Head weights were fine tuned for the first 20 epochs of training at a learning rate of 0.001 followed by fine tuning of all network weights at a learning rate of 0.0001 for a total of 200 training epochs. Trained model weights were saved after each epoch. The loss function decayed and average precision increased (figure S9) in both training and validation, indicating mask R-CNN learned iSCAT particle signals and showed no signs of overfitting. Visual inspection of the performance of mask R-CNN (figure S10) indicated mask R-CNN learned particle signals successfully. On a standard desktop computer (8 GB 2070 NVIDIA RTX GPU, AMD Ryzen 5 2600X 6 Core CPU, 64 MB RAM), training on the 500-image synthetic dataset was completed in ~40 minutes. Inference predictions of particle detection by mask R-CNN on experimental iSCAT data were inspected visually using weights from the 100^{th} and 200^{th} epoch. The 200^{th} epoch showed the best performance despite a plateau in precision during training and validation after ~50 epochs. We hypothesize this extended training beyond average precision saturation allowed for enhanced learning of the background, and thus justified the use of weights from the 200^{th} epoch in all subsequent analyses.

Single Particle Tracking

From PSFs detected in ratiometric images, the magnitude of the scattering contrast and event time for an adsorbing or desorbing particle was quantified using single particle tracking techniques. First, PSFs from adsorption events were detected using the Haar or mask R-CNN methods. For detections using the Haar method, bounding boxes of 15 pixels around the centroids of the proposed particle detections were generated. For detections using mask-RCNN, bounding boxes from inference were directly used. Any candidate particle with a bounding box center within 5 pixels of the image border was discarded. The PSFs of candidate particles were fit to a 2D-Gaussian[31] using a least-squares regression algorithm from the LMFIT[32] python package using

equation 4 where contrast amplitude, A, and center position (x_0, y_0) of the Gaussian were allowed to vary, while the background, B, and standard deviation, σ, were fixed at 1.0 contrast units and 2.0 pixels (~100 nm), respectively. Fixed values were determined by empirical fits to PSFs manually sampled from ratiometric images. In Haar detections, the signal to noise ratio, *SNR*, for each proposed detection was calculated as the ratio of $|A|$ to the background standard deviation and all detections with an SNR less than 3 were discarded. The A, frame, time, and position for all PSFs were logged. After PSF detection and fitting, PSFs with center positions within 2 pixels of each other in subsequent ratiometric images for a minimum of 6 frames were linked into trajectories using the Trackpy python package.[33] Within each trajectory, the average of the two highest magnitude amplitude ($|A|$) points was used to separate the data for a given particle trajectory into regions of opposite slope. Two lines were fit to A as a function of time in each region, and the position of intersection of these lines was defined as the relative scattering contrast, $|A_{event}|$, and event time at sub-frame interval temporal resolution. We note that for convenience $|A_{event}|$ will be referred to as scattering contrast for the remainder of this work. An example of the single particle tracking techniques used to quantify the scattering contrasts and event times of an adsorption event is illustrated in figure S8 (d) and (e).

RESULTS

Stage Vibrations Lead to False Positives

The LbL films used in the model experiment analyzed here exhibited higher roughness compared to minimally functionalized glass coverslips used in other iSCAT experiments (AFM images in figures S3 and S4, $R_{rms,LbL}$ = 3.2 nm, $R_{rms,glass}$ = 0.8 nm, *Area* = 10 μm x 10 μm). Figure 2 shows that when the LbL film roughness was coupled with stage movements, strong scattering features (red arrows, left) manifested as particle like features in ratiometric images (red arrows,

right). These background features had contrasts and morphologies similar to the AuNP scattering signals (blue arrows, right). The Haar method used to detect particles frequently identified these features as particles leading to false positive identifications. In this way, figure 2 highlights the limitations of using edge and threshold-based image analysis algorithms, such as the Haar method, to detect particles.

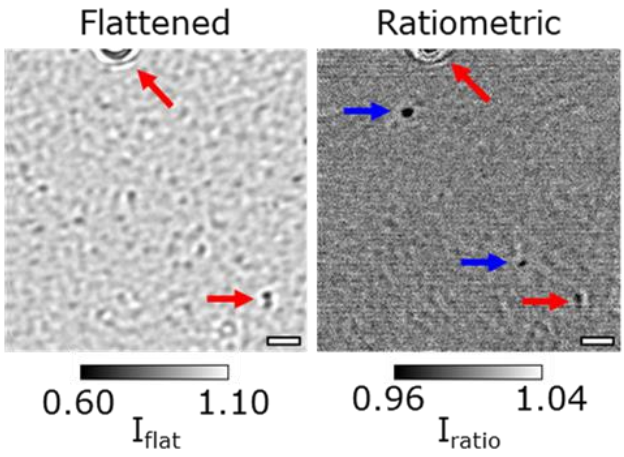


Figure 2. (left) A flattened image from an iSCAT experiment of 19.2 nm AuNPs adsorbing to LbL films with 3.2 nm roughness. Increased substrate roughness created regions of high scattering contrast, indicated by the red arrows. Scattering the background dominated the image and particle signatures were indistinguishable from the background. (right) A ratiometric image from the same experiment. Stage movements caused strongly scattering roughness features to show up the background of the ratiometric images as indicated by the red arrows. These background features have similar contrast and morphologies to the scattering features from AuNPs as indicated by the blue arrows. All scale bars are 1 μm.

Reducing False Positive Detections with Mask R-CNN

CNNs have been shown to be successful in object detection and to be less sensitive to algorithm parameters that normally must be tuned for object detection in every experiment.[21,22] To see if these benefits translate to iSCAT image processing, a hybrid dataset containing background

images from an iSCAT experiment and simulated particle PSFs was generated as described in the Methods section. This dataset was used to fine tune mask R-CNN network weights via transfer learning, giving a CNN adept at identifying and classifying common place objects (e.g., animals, balls) the ability to identify and classify nanoparticle PSFs in ratiometric iSCAT images. In figure 3, we compare the performance of particle detection using mask R-CNN with the Haar method by analyzing the data from the same AuNP-LbL adsorption experiment using both methods.

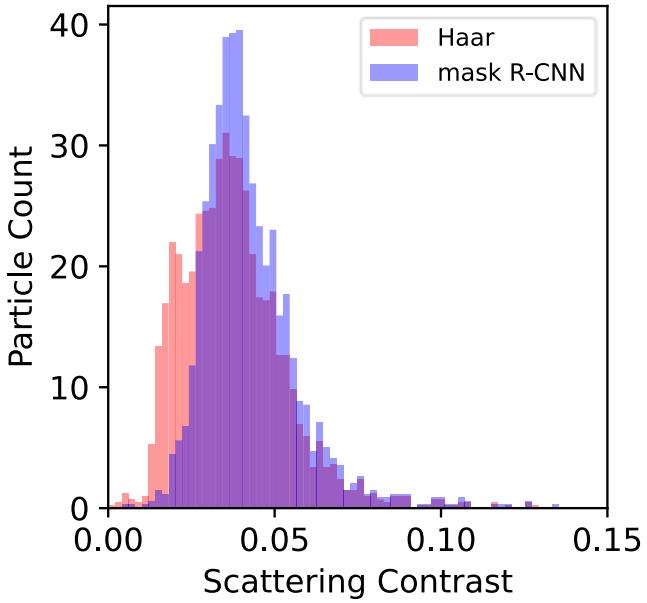


Figure 3. Histograms of scattering contrast for particles detected using the Haar method (pink) and mask R-CNN (blue). Scattering Contrast ($|A_{event}|$) represents the peak contrast amplitude for each particle as determined by the single particle tracking algorithm.

Figure 3 shows the distribution of scattering contrasts measured using ratiometric image processing and detected using the Haar method (pink) and mask R-CNN (blue). In the Mask R-CNN histogram, the distribution is monomodal with a peak at ~0.040. The Haar histogram is bimodal, with a dominant peak at ~0.04 and a second lower peak at ~0.02. (Peak position and count were determined by eye). The bimodal nature of scattering contrasts detected by the Haar method demonstrates the main improvement in performance of mask R-CNN over the Haar

method. Despite our particles being monodisperse in size (SEM image in figure S6), the Haar method detects multiple peaks in ratiometric scattering contrast. The low contrast peak near 0.02 in the Haar histogram, indicative of false positives during particle detection, is not present in the mask R-CNN histogram. Moreover, the breadth of the low scattering intensity peak near 0.02 in the Haar histogram overlaps significantly with the breadth of the central peak in the Haar histogram at 0.04. Because of this overlap, post-processing to remove false positive detections without excluding true positive detections was not feasible. In addition, Mask R-CNN's detection capabilities are invariant to parameter settings such as minimum SNR (figure S11**)** while the Haar algorithm is very sensitive to the minimum SNR value, leading to improved trajectory linking during particle tracking. This invariance also emphasizes the ability of mask R-CNN to identity particles in complex backgrounds at low SNR values. And though beyond the scope this work, the ability to detect objects at a low SNRs would extend the capability of iSCAT experiments to detect weaker scattering objects such as proteins or small, low dielectric nanoparticles in complex, heterogenous backgrounds as has been recently shown by Dahmardeh[34] *et al.* in a publication currently under review.

Learning the Background Improves Performance with Mask R-CNN

The ability of CNNs to detect objects has been attributed to the ability of the network to learn features related to the objects of interest.[22,35] However, CNNs also inherently learn to classify features as belonging to the background. The overlap of the dominant peaks at 0.04 in both histograms in figure 3 suggests that detecting particles is not the main challenge, but rather the problem lies in discerning particles from the background - the inverse problem in object detection. And because there is no bimodal behavior in the mask R-CNN histogram due to false positive detections at low scattering contrasts, the improved performance of mask R-CNN may not stem

from an ability to learn a better representation of the particle signals, but instead its ability to learn a better representation of the background in iSCAT ratiometric images. Therefore, the inclusion of representative background images that capture real experimental backgrounds in datasets for training may be important.

To test the hypothesis that including experimentally representative backgrounds in the dataset improves mask R-CNN's performance relative to edge and threshold based methods like the Haar-method, an analogous dataset was generated without the inclusion of representative background images from ratiometric images. Figure 4 presents histograms of the scattering contrast detected by mask R-CNN networks trained with (blue) and without (gray) experimental backgrounds in the dataset images.

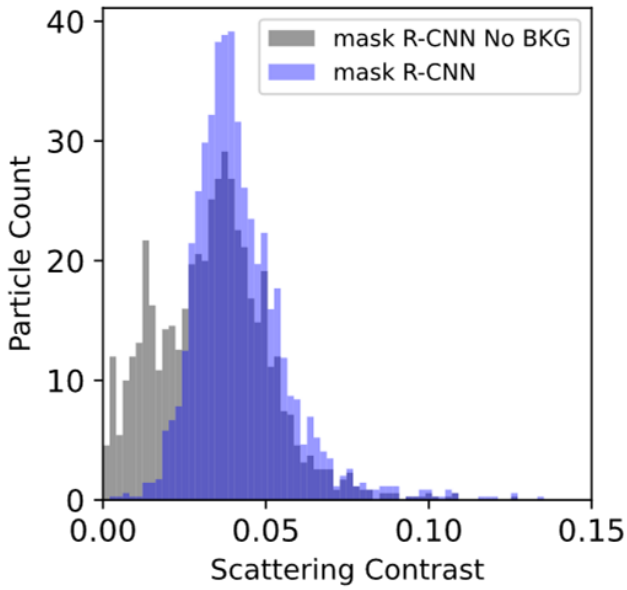


Figure 4. Histograms of scattering contrast comparing the performance of mask R-CNN models trained on datasets that included (blue) and did not include (gray) experimental backgrounds. Scattering Contrast ($|A_{event}|$) represents the peak contrast amplitude for each particle as determined by the single particle tracking algorithm.

Figure 4 demonstrates the impact of including representative experimental backgrounds in the training dataset. An increased density of false positive detections appears at scattering contrast values centered around 0.01 when particle detection was executed with the mask R-CNN network trained on a dataset that did not include information about the background. The central scattering peak at 0.04 was also slightly decreased in particle count when data was processed using mask R-CNN trained on the dataset that did not contain experimental backgrounds. (Peak position and count were again determined by eye). This indicates that, in addition to reducing false positives, insufficient learning of background features may result in missed detections. Consequently, the improved performance of mask R-CNN is not solely a product of learning of the background features. Rather, it is a combination of the ability of mask R-CNN to learn the background and to better detect particle features in that background when the mask R-CNN is exposed to experimental backgrounds during training.

DISCUSSION

Visual Inspection of 100 Randomly Selected Images

To quantify the performance of the Haar method against both mask R-CNN models (trained using datasets with and without background), 100 random images were selected from the experimental ratiometric images. For each image, the three object detection methods described in the Results section were used to detect particles in the same sampled image, and the results were inspected visually to assess performance by counting the number of true positive (TP), false positive (FP), and false negative (FN) detections. Examples of true and false positive detections are shown in figure 2. False negatives are particle signals that have not been detected by the algorithm as potential PSFs. From this, precision and recall were calculated, and the results are tabulated in table 2. We note that true negative detections cannot be quantified in object detection

tasks as there are an infinite number pixel groupings that form background features that should not be detected.[36] Due to the low concentration of particles in solution, landing events occurred infrequently during our experiment, resulting in ratiometric images that contained 0-2 particles on average.

Table 2. Summary of results of particle detection in 100 randomly selected iSCAT images analyzed by visual inspection across the three methods tested. [a]Precision is defined as TP/(TP + FP) and [b]Recall is defined as TP/(TP + FN).

	Haar	**Mask R-CNN (no background)**	**Mask R-CNN (with background)**
Total Particles (Ground Truth)	81	81	81
True Positive Count (TP)	77	68	79
False Positive Count (FP)	19	21	3
False Negative Count (FN)	4	13	2
Precision[a]	80.2%	76.4%	96.3%
Recall[b]	95.1%	84.0%	97.5%

Table 2 summarizes the results of the visual inspection of 100 random images. The biggest difference in performance across the three methods is seen in the FP count where the Haar, Mask R-CNN (no background), and Mask R-CNN methods incorrectly identified 19, 21, and 3 background features as particle PSFs. In general, the numbers in table 2 agree with the distributions of scattering contrast shown in figures 3 and 4 and support the observation that the secondary, low scattering contrast peaks at 0.010 represent false positives and have the biggest impact on the quality of PSF detection in these iSCAT images.

In table 2, we also quantified performance using precision and recall metrics. Recall quantifies the ability of each method to capture all possible PSF-like features without missing features that represent true particle PSFs. Precision quantifies the accuracy of those detections, and thus quantifies the impact of FPs on the analysis. An optimal particle detection algorithm would

detect all features that represent particle PSFs (high recall) without incorrectly identifying background features as particle PSFs (high precision). The Haar method recall is 95.1% indicating that the Haar method detects particles well, but precision is 80.2% as many background features are identified as particle signals. In the mask R-CNN (no background) method, both recall (76.4 %) and precision (84.0%) are lower than the Haar Method indicating that, without the background in the training datasets, the neural network suffers from falsely detecting background features as particle signals and from missing real particle signals when they are convoluted with the background. We see the best performance in Mask R-CNN (with background) where high recall (97.5%) and precision (96.3%) show that Mask R-CNN, when trained using a dataset that includes experimental backgrounds, correctly identifies particle PSFs while also correctly avoiding classifying background features particles. Evidently, neural networks demonstrate the potential to learn experimental backgrounds, and this translates to improved performance in particle detection in iSCAT experiments.

False Positive Detections Negatively Impact Single Particle Tracking

The analysis described in the Methods section utilized single particle tracking algorithms to link PSFs detected into trajectories based on the proximity of the PSF centroid locations in subsequent ratiometric images. We then used these trajectories to fit lines to the PSF amplitude versus time. When a feature representing an adsorbing particle in a ratiometric video was correctly identified, the trend of scattering contrast with time within a trajectory followed a linear decrease to the time of the adsorption event followed by a linear increase after the adsorption event as shown in the example in figure S8 (e). However, when false positive detections of the background features are present, particle tracking algorithms may link a true particle detection with a false positive if the false position is close to the true positive position from the previous image. When this happens,

the trends in A with time for a given trajectory can deviate from the expected decreasing-increasing pattern and this causes the linear fit algorithm to either fail or report an incorrect scattering contrast and adsorption event time. In general particle tracking experiments, such as those that measure diffusion, incorrect linking can skew the detected velocity.[37] Thus, the improvements in particle detection demonstrated here could benefit all particle tracking studies, and our dataset creation/training methodology provides a single particle tracking strategy with improved reliability.

Utility Beyond iSCAT

The main contribution here is the methodology behind including real experimental background images in datasets for training CNNs via supervised machine learning. Including real background images in the training datasets, leads to an improvement in particle detection and, more importantly, a reduction in false positives when iSCAT images were analyzed with Mask R-CNN. However, the process developed here is not limited to iSCAT experiments. If experimentalists can accurately capture the background in their collected images and model signals (e.g., PSFs) they wish to detect, they can create labeled datasets for many types of experimental images (e.g., atomic force microscopy, transmission and scanning electron microscopy, other optical microcopy methods) to aid in the development of accurate, high-throughput processing of their data using machine learning without the need for expensive, high-performance computing resources.

A particular challenge for further expansion of the methods for creating labeled datasets that include experimental backgrounds described here is that it requires a researcher to have access to background images lacking the target objects of interest. Dataset creation in this work was enabled by the low (200 pM) concentration nanoparticle solutions used in our adsorption

experiment. The low concentration of nanoparticles slowed the adsorption process relative to the frame rate of iSCAT (184 fps) to the point that only 0-2 events were detected per frame, and resulted in multiple ratiometric images that contained no particle signatures. Furthermore, for characterization tools compatible with solutions such as iSCAT, a blank solution could also have been imaged without particles. This situation, though, is inefficient from a data versus information standpoint and is uncommon in other characterization methods. To implement the dataset creation method in experiments where all images contain objects of interest, users would need to remove the objects from their images to create signal free background images, which requires detecting them, rendering the workflow obsolete. A promising approach to extend the method of creating datasets with real experimental backgrounds may exist in the computer vision machine learning field in the form of neural style transfer.[38] In neural style transfer, a neural network takes two images as inputs, a content image and a style reference, and blends the two together such that the output image contains important features in the content image "painted" in the style of the reference image. In our workflow, the content image would be the particle image in figure 1 and the reference would be sets of images from an experiment with or without particles (i.e., the experimental ratiometric background image in figure 1). In theory, the output from this network could result in a dataset containing synthetic particles in an image atop a realistic background with labels from the particle image serving as ground truth labels for training.

CONCLUSIONS

The methods presented here demonstrate a workflow for creating labeled datasets with representative experimental backgrounds containing simulated particle signals on background suppressed interferometric scattering (iSCAT) microscopy images to use in training a mask region based convolutional neural network (mask R-CNN) using transfer learning without extensive

computational resources. Results showed that the high performance of mask R-CNN in detecting nanoparticles adsorbing to a rough, high scattering layer-by-layer film in an iSCAT experiment stemmed from the ability of mask R-CNN to learn background features, thereby reducing the number of false positive detections. The improved analysis technique expands capability of iSCAT to detect scattering objects in situations where background scattering is complex.

ACKNOWLDEGEMENT

We thank the National Science Foundation (NSF) Graduate Research Fellowships Program (GRFP), NSF Chemical, Bioengineering, Environmental, and Transport Systems (CBET) 2034122, National Institutes of Health (NIH) R35GM118139, NSF Division of Materials Research (DMR) Polymers Program DMR1905912, NSF Partnerships for International Research, Center for Engineering MechanoBiology NSF Science and Technology Center, CMMI: 15-48571, and the Education for the Office of International Science and Engineering PIRE-OISE-1545884 that provided support for this work. This work also was carried out in part at the Singh Center for Nanotechnology, which is supported by the NSF National Nanotechnology Coordinated Infrastructure Program under grant NNCI-2025608. This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, IM LLNL-JRNL-842942. Lastly, we thank Drs. Nathan Frey, James Horwath, and Chris Price for discussions regarding the implemented machine learning methods, Gautam Jain for AFM images, Drs. Daeyeon Lee and Wilfredo Mendez Ortiz for assistance in preparing LbL films, Dr. Matthew Capporizzo for assistance in operating the iSCAT equipment, and Drs. Jan Genzer and Yeongun Ko at North Carolina State University for fruitful discussions.

ASSOCIATED CONTENT

Supporting information contains details regarding image flattening, ratiometric processing, mask R-CNN training, iSCAT instrumentation and flow cells, materials, substrate preparation, substrate characterization, and gold nanoparticle characterization.

REFERENCES

(1) Taylor, R. W.; Sandoghdar, V. Interferometric Scattering (ISCAT) Microscopy and Related Techniques. In *Label-Free Super-Resolution Microscopy*; Astratov, V., Ed.; Springer International Publishing: Cham, **2019**; pp 25–65.DOI: 10.1007/978-3-030-21722-8_2.

(2) Strack, R. Scattering Microscopy Takes Single-Particle Tracking to the next Level. *Nat. Methods* **2019**, *16* (6), 455. DOI: 10.1038/s41592-019-0438-3.

(3) Holanová, K.; Vala, M.; Piliarik, M. Optical Imaging and Localization of Prospective Scattering Labels Smaller than a Single Protein. *Opt. Laser Technol.* **2019**, *109* (August 2018), 323–327. DOI: 10.1016/j.optlastec.2018.08.014.

(4) Jacobsen, V.; Stoller, P.; Brunner, C.; Vogel, V.; Sandoghdar, V. Interferometric Optical Detection and Tracking of Very Small Gold Nanoparticles at a Water-Glass Interface. *Opt. Express* **2006**, *14* (1), 405. DOI: 10.1364/opex.14.000405.

(5) Spindler, S. *et al.* Visualization of Lipids and Proteins at High Spatial and Temporal Resolution via Interferometric Scattering (ISCAT) Microscopy. *J. Phys. D. Appl. Phys.* **2016**, *49* (27). DOI: 10.1088/0022-3727/49/27/274002.

(6) Kukura, P.; Ewers, H.; Müller, C.; Renn, A.; Helenius, A.; Sandoghdar, V. High-Speed Nanoscopic Tracking of the Position and Orientation of a Single Virus. *Nat. Methods* **2009**, *6* (12), 923–927. DOI: 10.1038/nmeth.1395.

(7) Heermann, T.; Steiert, F.; Ramm, B.; Hundt, N.; Schwille, P. Mass-Sensitive Particle Tracking to Elucidate the Membrane-Associated MinDE Reaction Cycle. *Nat. Methods* **2021**, *18* (10), 1239–1246. DOI: 10.1038/s41592-021-01260-x.

(8) Foley, E. D. B.; Kushwah, M. S.; Young, G.; Kukura, P. Mass Photometry Enables Label-

Free Tracking and Mass Measurement of Single Proteins on Lipid Bilayers. *Nat. Methods* **2021**, *18* (10), 1247–1252. DOI: 10.1038/s41592-021-01261-w.

(9) Ortega Arroyo, J.; Andrecka, J.; Spillane, K. M.; Billington, N.; Takagi, Y.; Sellers, J. R.; Kukura, P. Label-Free, All-Optical Detection, Imaging, and Tracking of a Single Protein. *Nano Lett.* **2014**, *14* (4), 2065–2070.DOI: 10.1021/nl500234t.

(10) Andrecka, J.; Takagi, Y.; Mickolajczyk, K. J.; Lippert, L. G.; Sellers, J. R.; Hancock, W. O.; Goldman, Y. E.; Kukura, P. Interferometric Scattering Microscopy for the Study of Molecular Motors, in *Methods in Ezymology*, 1st ed.; Elsevier Inc., **2016**; Vol. 581. DOI: 10.1016/bs.mie.2016.08.016.

(11) Young, G.; *et al.* Quantitative Mass Imaging of Single Biological Macromolecules. *Science (80-.).* **2018**, *360* (6387), 423–427. DOI: 10.1126/science.aar5839.

(12) Cole, D.; Young, G.; Weigel, A.; Sebesta, A.; Kukura, P. Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy. *ACS Photonics* **2017**, *4* (2), 211–216. DOI: 10.1021/acsphotonics.6b00912.

(13) Melo, L.; Hui, A.; Kowal, M.; Boateng, E.; Poursorkh, Z.; Rocheron, E.; Wong, J.; Christy, A.; Grant, E. Size Distributions of Gold Nanoparticles in Solution Measured by Single-Particle Mass Photometry. *J. Phys. Chem. B* **2021**, *125* (45), 12466–12475. DOI: 10.1021/acs.jpcb.1c05557.

(14) Ortega Arroyo, J.; Cole, D.; Kukura, P. Interferometric Scattering Microscopy and Its Combination with Single-Molecule Fluorescence Imaging. *Nat. Protoc.* **2016**, *11* (4), 617–633. DOI: 10.1038/nprot.2016.022.

(15) Taylor, R. W.; Sandoghdar, V. Interferometric Scattering Microscopy: Seeing Single Nanoparticles and Molecules via Rayleigh Scattering. *Nano Lett.* **2019**, *19* (8), 4827–

4835. DOI: acs.nanolett.9b01822.

(16) Strack, R. ISCAT Gets a Signal Boost. *Nat. Methods* **2021**, *18* (5), 444–448. DOI: 10.1038/s41592-021-01157-9.

(17) Liebel, M.; Hugall, J. T.; Van Hulst, N. F. Ultrasensitive Label-Free Nanosensing and High-Speed Tracking of Single Proteins. *Nano Lett.* **2017**, *17* (2), 1277–1281. DOI: 10.1021/acs.nanolett.6b05040.

(18) Cheng, C. Y.; Hsieh, C. L. Background Estimation and Correction for High-Precision Localization Microscopy. *ACS Photonics* **2017**, *4* (7), 1730–1739. DOI: 10.1021/acsphotonics.7b00238.

(19) Chatzidakis, M.; Botton, G. A. Towards Calibration-Invariant Spectroscopy Using Deep Learning. *Sci. Rep.* **2019**, *9* (1), 1–10. DOI: 10.1038/s41598-019-38482-1.

(20) He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42* (2), 386–397. DOI: 10.1109/TPAMI.2018.2844175.

(21) Horwath, J. P.; Zakharov, D. N.; Mégret, R.; Stach, E. A. Understanding Important Features of Deep Learning Models for Segmentation of High-Resolution Transmission Electron Microscopy Images. *npj Comput. Mater.* **2020**, *6* (1), 1–9. DOI: 10.1038/s41524-020-00363-x.

(22) Frei, M. .; Kruis, F. E. Image-Based Analysis of Dense Particle Mixtures via Mask R-CNN. *Eng* **2022**, *3* (1), 78–98. DOI: 10.3390/eng3010007.

(23) Abdulla, W. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. *GitHub repository*. Github **2017**.

(24) Panigrahi, S.; Nanda, A.; Swarnkar, T. A Survey on Transfer Learning. *Smart Innov. Syst. Technol.* **2021**, *194* (10), 781–789. DOI: 10.1007/978-981-15-5971-6_83.

(25) Yang, L.; Parton, R.; Ball, G.; Qiu, Z.; Greenaway, A. H.; Davis, I.; Lu, W. An Adaptive Non-Local Means Filter for Denoising Live-Cell Images and Improving Particle Detection. *J. Struct. Biol.* **2010**, *172* (3), 233–243. DOI: 10.1016/j.jsb.2010.06.019.

(26) Young, G.; Kukura, P. Interferometric Scattering Microscopy. *Annu. Rev. Phys. Chem.* **2019**, *70* (1). DOI: 10.1146/annurev-physchem-050317-021247.

(27) Ortega-Arroyo, J.; Kukura, P. Interferometric Scattering Microscopy (ISCAT): New Frontiers in Ultrafast and Ultrasensitive Optical Microscopy. *Phys. Chem. Chem. Phys.* **2012**, *14* (45), 15625–15636. DOI: 10.1039/c2cp41013c.

(28) Ortega Arroyo, J.; Cole, D.; Kukura, P. Interferometric Scattering Microscopy and Its Combination with Single-Molecule Fluorescence Imaging. *Nat. Protoc.* **2016**, *11* (4), 617–633. DOI: 10.1038/nprot.2016.022.

(29) Sources of vibration. https://www.thorlabs.com/tutorials/Tables2.cfm. (accessed June 28, 2022).

(30) Piliarik, M.; Vala, M. Weighing Single Protein Complexes on the Go. *Nat. Methods* **2021**, *18* (10), 1155–1156. DOI: 10.1038/s41592-021-01259-4.

(31) Shen, H.; Tauzin, L. J.; Baiyasi, R.; Wang, W.; Moringo, N.; Shuang, B.; Landes, C. F. Single Particle Tracking: From Theory to Biophysical Applications. **2017**. DOI: 10.1021/acs.chemrev.6b00815.

(32) Newville, M.; Stensitzki, T.; Allen, D. B.; Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python (0.8.0). Zenodo **2014**. DOI: DOI: 10.5281/zenodo.11813.

(33) Allan, D. *et al.* Soft-Matter/Trackpy: Trackpy v0.4.2. Zenodo October **2019**. DOI: 10.5281/zenodo.3492186.

(34) Dahmardeh, M.; Dastjerdi, H. M.; Mazal, H.; Köstler, H.; Sandoghdar, V. Self-Supervised Machine Learning Pushes the Sensitivity Limit in Label-Free Detection of Single Proteins below 10 KDa. *Research Square* **2022**. DOI: 10.21203/rs.3.rs-1635524/v1.

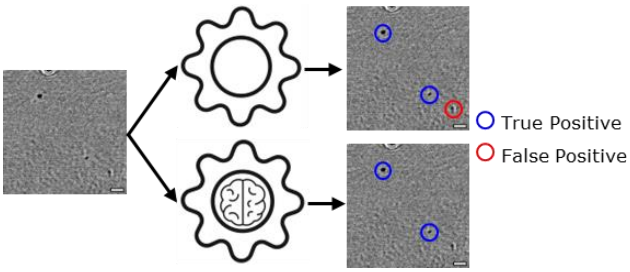
(35) Newby, J. M.; Schaefer, A. M.; Lee, P. T.; Forest, M. G.; Lai, S. K. Convolutional Neural Networks Automate Detection for Tracking of Submicron-Scale Particles in 2D and 3D. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (36), 9026–9031. DOI: 10.1073/pnas.1804420115.

(36) Padilla, R.; Netto, S. L.; da Silva, E. A. B. A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*; **2020**; pp 237–242. DOI: 10.1109/IWSSIP48289.2020.9145130.

(37) Rose, K. A.; Molaei, M.; Boyle, M. J.; Lee, D.; Crocker, J. C.; Composto, R. J. Particle Tracking of Nanoparticles in Soft Matter. *J. Appl. Phys.* **2020**, *127* (19), 191101. DOI: 10.1063/5.0003322.

(38) Gatys, L.; Ecker, A.; Bethge, M. A Neural Algorithm of Artistic Style. *J. Vis.* **2016**, *16* (12), 326. DOI: /10.1167/16.12.326.

TABLE OF CONTENTS IMAGE



Supporting Information for Enhancing Nanoparticle Detection in Interferometric Scattering (iSCAT) Microscopy Using Mask R-CNN

Michael J. Boyle[1,2] (boyle17@llnl.gov), Yale E. Goldman[3] (goldmany@upenn.edu) and Russell J. Composto[1] * (composto@seas.upenn.edu)

[1]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA, 19104, United States.

[2]Center for Engineered Materials and Manufacturing, Lawrence Livermore National Laboratory, Livermore, CA, 94550, United States

[3]Department of Physiology and Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, PA, 19104, United States.

Section S1: Materials

Poly(allylamine) hydrochloride (PAH, average M_w = 50 kg/mol, 283223-5G), 35% w./w. solution of poly(acrylic acid) in H_2O (PAA, average M_w = 250 kg/mol, 416002), sodium phosphate dibasic heptahydrate (Na_2PO_4H, S9390), Trizma hydrochloride (TRIS, T5941), sodium hydroxide (NaOH, 221465), 37% w./w. hydrochloric acid (HCl, 258148), sodium dodecyl sulfate (SDS, 436143) were purchased from Sigma Aldrich and used for buffer preparation, layer-by-layer (LbL) film deposition, and QCM-D flow cell cleaning. Sulfuric acid (H_2SO_4, 98%, A300SI-212) and hydrogen peroxide (H_2O_2, H325-500) were purchased from Fisher Scientific for piranha etching. All chemicals were used as purchased and all solutions were prepared using Milli-Q deionized water. Custom 20 nm gold nanoparticles with carboxyl polyethylene glycol (lipoic acid-NH-PEG_{12}-COOH) ligands in deionized water (c_{stock} = 16.25 nM) were purchased from nanoComposix, diluted to 200 pM in buffer solution for experiments, and filtered to remove dust and contaminants from the solutions using 0.2 µm PTFE syringe filters (Sigma Aldrich, SLFG025NS) prior to experiments.

Section S2: Substrate Preparation

Layer-by-Layer Film Deposition

No. 1.5 (22 mm x 40 mm x 150 µm, 12-544-B, Fisher Scientific) glass coverslips and silicon wafers were immersed in piranha etching solution (45 mL 98% H_2SO_4:15 mL 30% H_2O_2) at 80°C for 30 minutes. Substrates were rinsed 3 times and stored in water overnight. Immediately before LbL deposition, substrates were dried under streaming N_2 gas. After drying, substrates were exposed to UV-ozone (ProCleaner Plus, BioForce NanoSciences) for 10 minutes. SiO_2 coated quartz QCM-D crystals were purchased from Nanoscience (BL-QSX 303). Sensors were immersed in piranha etching solution (15 mL 98% H_2SO_4:5 mL 30% H_2O_2) at room temperature

for 10 seconds. Sensors were rinsed 3 times with water and dried under streaming N_2 gas. Immediately before LbL deposition, sensors were exposed to UV-ozone for 10 minutes.

LbL films were prepared following procedures detailed by Yoo[1] *et al*. Poly(allylamine hydrochloride) (PAH) with a molecular weight of 50 kg/mol and poly(acrylic acid) (PAA) of molecular weight 250 kg/mol from Sigma Aldrich served as the polycation and polyanion respectively. PAH was purchase in salt form and diluted in acidic water (pH = 3.0) to obtain a 0.01 M monomer concentration solution. PAA was purchased in a 35 wt.% solution in water and diluted with DI water to a final monomer concentration of 0.01 M, giving a 1:1 ratio of cation to anion groups. After ample mixing, the pH of the PAA solution was reduced to 3.5 with a 1 M HCl solution, and the pH of the PAH solution was increased to 5.6 using 1 M NaOH. Multilayer films were assembled on cleaned silicon wafer, coverslips, and QCM-D sensors by using an automatic dipping method with a MICROM DS 50 slide stainer in batches of six. For one bilayer, the dipping procedure comprised four sequential steps of PAH deposition for 15 min, and three rinsing steps with water of 2 min, 2 min, and 1 min, respectively, followed by four identical steps for PAA deposition. In total 5.5 bilayers were assembled resulting in positively charged films with a PAH cap. After coating, the films were rinsed in DI water. Deposition uniformity and quality were confirmed by visual inspection.

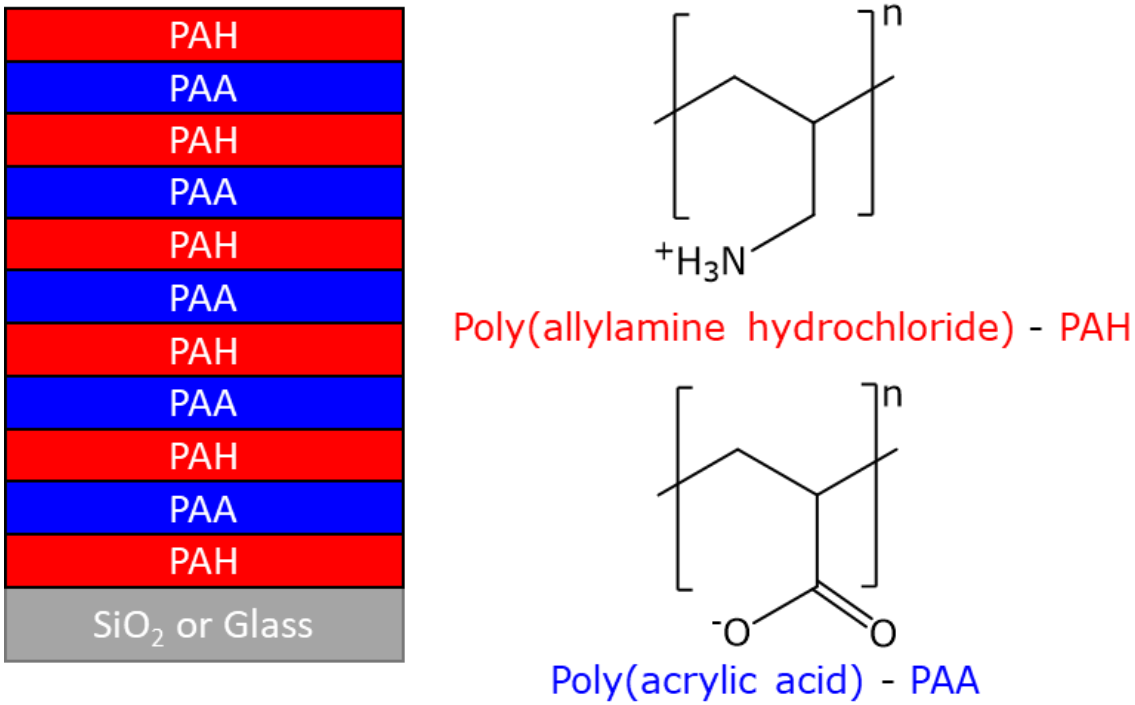


Figure S1. (left) LbL film structure after deposition. Layers are drawn as discrete entities based on the alternating dip coating process described above, though it has been reported that layers interpenetrate and boundaries are diffuse in these systems.[1] (right) Chemical structures of the polycation (PAH) and polyanion (PAA) used in LbL deposition.

Layer-by-Layer Film Characterization

The results of LbL film characterization are listed in table S1 Methods for determining values listed are detailed below.

Table S1. Summary of LbL film properties.

Bilayers	**Layers**	**Cap**	$t_{LbL,dry}$	$t_{LbL,wet}$	ζ_{LbL}	**WCA**	$R_{Q,LbL}$	$R_{Q,glass}$
5.5	11	PAH	60 nm	85 nm	+14.7 mV	16.4°	3.2 nm	0.8 nm

Thickness ($t_{LbL,dry}$ and $t_{LbL,wet}$) Measured via Ellipsometry

All ellipsometry measurements were carried out using an Alpha-SE Ellipsometer from J.A. Woolam. Dry thickness ($t_{LbL,dry}$) measurements of LbL films were fit to a Cauchy model where the

refractive index was allowed to vary between 1.5 and 1.6, with an ambient refractive index of 1 representing air.

Figure S2. (a) Liquid cell for wet ellipsometry. (b) and (c) Snapshots showing configuration of liquid cell during measurements.

For wet ellipsometry thickness ($t_{LbL,wet}$) measurements, a homemade, 3D-printed liquid cell (figure S2) was used where the incidence angle of light on the cell was designed to match the 70° incidence angle of the device. Dry thickness was measured in the liquid cell, followed by filling of the cell with 1 mM Na_2PO_4H buffer at a pH of 7.0. Wet thickness was measured as a function of swelling time, reaching equilibrium after 1-2 minutes of buffer exposure. Raw data was fit to a Cauchy model where thickness and refractive index was allowed to vary between 1.4 and 1.5, with an ambient refractive index for water of 1.33. Results reported represent an average of three measurements after 10 minutes of buffer exposure.

Zeta Potential (ζ_{LbL})

Following procedures detailed in literature,[2] the zeta potential of flat surface samples was measured with a DelsaNano-C (Beckmann Coulter) instrument using the flat surface cell. The electrophoretic mobility of 300 nm SiO_2 probe particles (NanoComposix, SISN300-25M) between the sample and cell surfaces was measured using electrophoretic light scattering. The zeta potential in volts (ζ) was calculated using the Smoluchowski Eqaution (equation S1):

$$\zeta_{LbL} = \frac{k\pi\eta U}{\varepsilon} \qquad \text{(S1)}$$

where k is the Smoluchowski coefficient, η is the viscosity of the solution at 25°C, ε is the dielectric constant of the solution at 25°C, and U is the electrophoretic mobility of probe particles. Each measurement was the average of 70 individual measurements performed at various positions (10 measurements each at 7 positions). All measurements were done in at least triplicate. Solution pH was measured with a dual pH/conductivity meter (Hanna Edge Dedicated pH/ORP meter). Solutions of 1 mM TRIS solutions with probe particles (140 μL, base frequency ~120 Hz) were prepared. Solution pH was regulated by titration with 1 M NaOH and 1 M HCl.

Water Contact Angle

Sessile drop water contact angles were measured using a home built static contact angle apparatus. To minimize swelling during measurements and capture wetting properties of films in experimental conditions, films were immersed in buffer solution (1 mM Na_2PO_4H buffer at a pH of 7.0) for 10 minutes and then dried under a stream of N_2 immediately prior to measurements. Contact angle profiles were measured using the LB-ADSA plugin in ImageJ.[3] Water contact angle (WCA) results represent the average of at least 3 independent measurements at different regions on the same sample.

Roughness ($R_{rms,LbL}$ and $R_{rms,glass}$) Measurements Using AFM

Surface profile measurements were carried out using an Asylum MFP-3D AFM with tapping mode tips (SCOUT 350 RAI, radius of curvature < 10 nm, NuNano). Wet measurements were conducted in tapping mode in an open fluid cell after LbL samples were exposed to 1 mM Na_2PO_4H buffer at a pH of 7.0 for 10 minutes to ensure equilibrium swelling conditions were reached, as confirmed via wet ellipsometry measurements. All AFM images were processed (plane fitting followed by row alignment via median of differences) using Gwyddion software. Roughness was also calculated using Gwyddion software. Results are shown in figure S3.

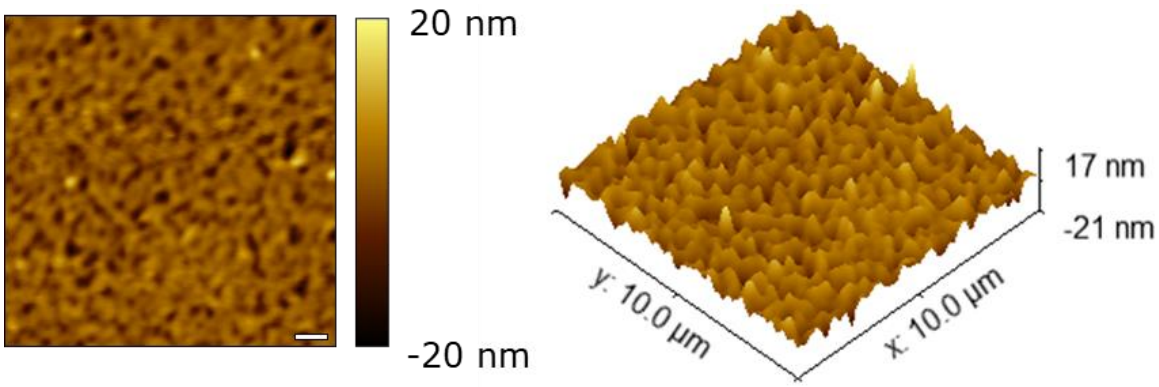


Figure S3. (left) A wet AFM topography image of a LbL film in 1 mM Na_2PO_4H buffer at a pH of 7.0. Scale bar is 1 µm. (right) An isometric 3D projection of the AFM image show to the left highlighting the rough nature of LbL films.

As a baseline, AFM topography measurements of neat coverslips (after piranha etching and UVO) were also collect were also collected and are shown in figure S4.

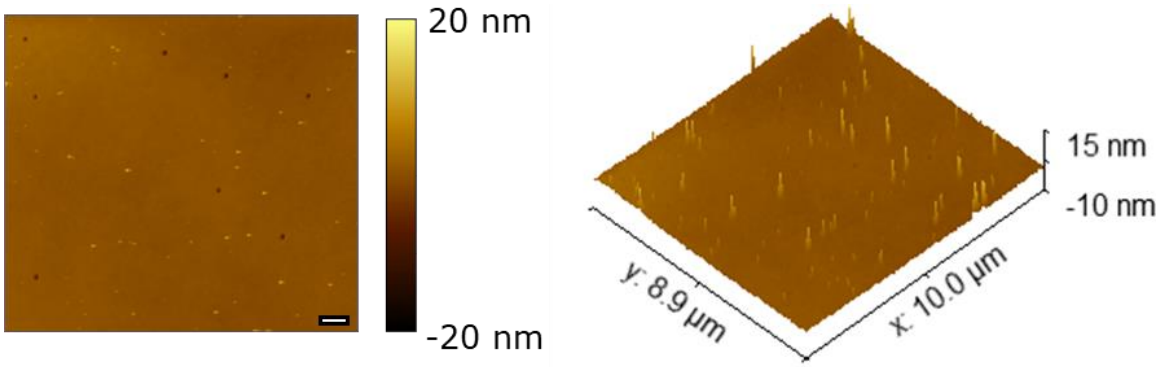


Figure S4. (left) An AFM topography image of a glass coverslip. Scale bar is 1 µm. (right) An isometric 3D projection of the AFM image left highlighting the flat nature of the surface relative to LbL films.

Section S3: Interferometric Scattering Microscopy Flow Experiments

Custom CoverWell perfusion chambers with adhesive added by the manufacturer were used as purchased (32 mm x 3 mm x 0.75 mm, Grace Bio Labs, dimensions in figure S5). LbL functionalized coverslips were attached to the chambers via the adhesive, forming leak-free,

reproducible geometry flow cells for iSCAT experiments. Polyethylene microfluidic tubing (BB61395, Scientific Commodities) connected the chambers and a microfluidic pump (Harvard PhD 2000) using press fit tubing connectors (460003, Grace Bio Labs) and a 20 mL syringe. To Facilitating long experiments (up to 10 hours) and reduce vibrations from fluid flow, the flow rate in the chamber was set and maintained at a low value measured to be 10 µL/min.

Experiments were carried out as follows. First, flow cell chambers were assembled on the iSCAT stage as shown in figure S5.

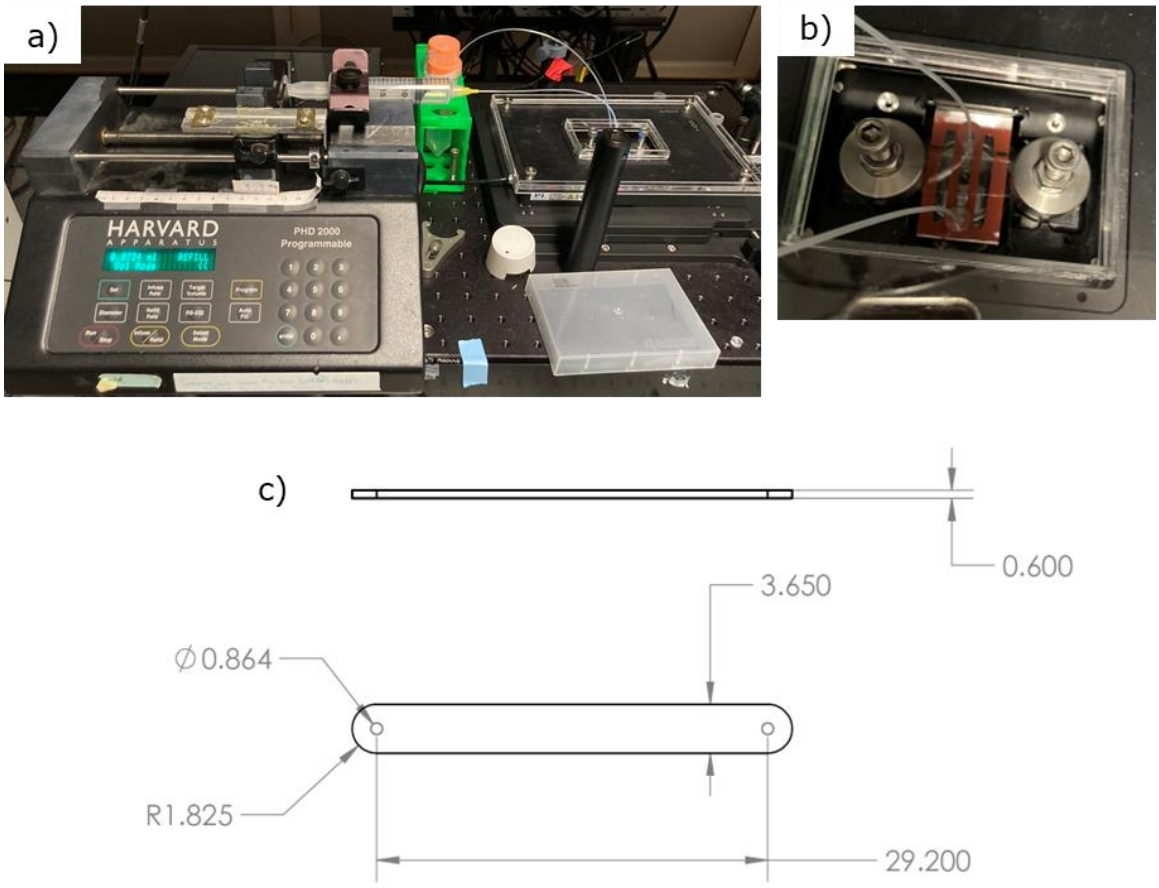


Figure S5. (a) An iSCAT flow cell mounted on the piezoelectric stage. The microfluidic pump generated negative pressure in the flow lines, pulling solution from the reservoir (a 50 mL centrifuge tube) through the flow cell at a constant rate. (b) A close-up image of the flow cell over

the objective lens shows the press fit tubing connectors and tubing at the inlet and outlet. (c) Dimensions of the flow cell. All dimensions are mm.

Next, sodium phosphate buffer solution was pumped from the reservoir across the objective and the time of travel of the liquid to the objective ($t_{to\ obj}$) was noted. LbL films were allowed to reach equilibrium in the flow cell in buffer for 10 minutes (equilibration time was determined by wet ellipsometry). 66 seconds of control iSCAT videos were taken of LbL films after swelling in buffer solution with no particles. Next, the reservoir solution was switched to a buffer solution containing gold nanoparticles at a 200 pM concentration. The pump was started and video capture was initiated after a period equal to $t_{to\ obj}$, the time when AuNP solution first reached the objective. 66 seconds of iSCAT videos were then collected every 30 minutes to analyze adsorption rates at varying levels of particle surface coverage.

Section S4: Gold Nanoparticles

The solution and particle properties of gold nanoparticles (AuNPs) with lipoic acid-PEG_{12}-COOH ligands in 1 mM Na_2PO_4H buffer at a pH of 7.0 are listed in table S2. Methods for determining values listed are detailed below.

Table S2. Summary of AuNP solution conditions and properties.

[Na_2PO_4H]	**[Au NP]**	**pH**	d_{SEM}	$d_{H,DLS}$	ζ_{NP}	κ_{debye}^{-1}
1 mM	200 pM	7.0 ± 0.1	18.5 ± 2.2 nm	20.1 ± 4.9 nm	-58 mV	5.56 nm

Solution Preparation

To stabilize pH during experiments, 1 mM buffer solutions of Na_2PO_4H in water were prepared. 1 M HCL was added to titrate solutions to a pH of 7.0. pH was verified using a Hanna Edge Dedicated pH/ORP meter. 0.256 mL of 16.25 nM AuNP stock solutions were added to 19.744 mL of buffer solution to yield 20 mL of 200 pM AuNP solutions in 1 mM Na_2PO_4H at a pH of 7.0 ± 0.1. pH was tested after addition of NPs, and no changes to pH were noted. All solutions were filtered to remove dust and contaminants from the solutions using 0.2 μm PTFE syringe filters prior to experiments.

Zeta Potential (ζ_{NP})

Zeta potential was measured by the supplier (NanoComposix) at similar experimental conditions.

Measuring Hard Sphere Diameter (d_{SEM}) with SEM

AuNPs were deposited on a LbL coated QCM-D sensor during a separate adsorption experiment. Sensors were removed from the flow cell and immediately dried under a stream of N_2. Sensors were then mounted on measurement scaffolds for a FEI Quanta 600 ESEM and imaged at a relative humidity of 0.38 bar of water vapor pressure and 30.00 kV acceleration in the secondary electron (SE) imaging mode at 0° tilt, providing a magnification of 208,203. Images were

segmented to determine particle size using a watershed transformation via custom python scripts available in the skimage package.[4] A representative SEM image and histogram describing the size of the nanoparticles are shown in figure S6.

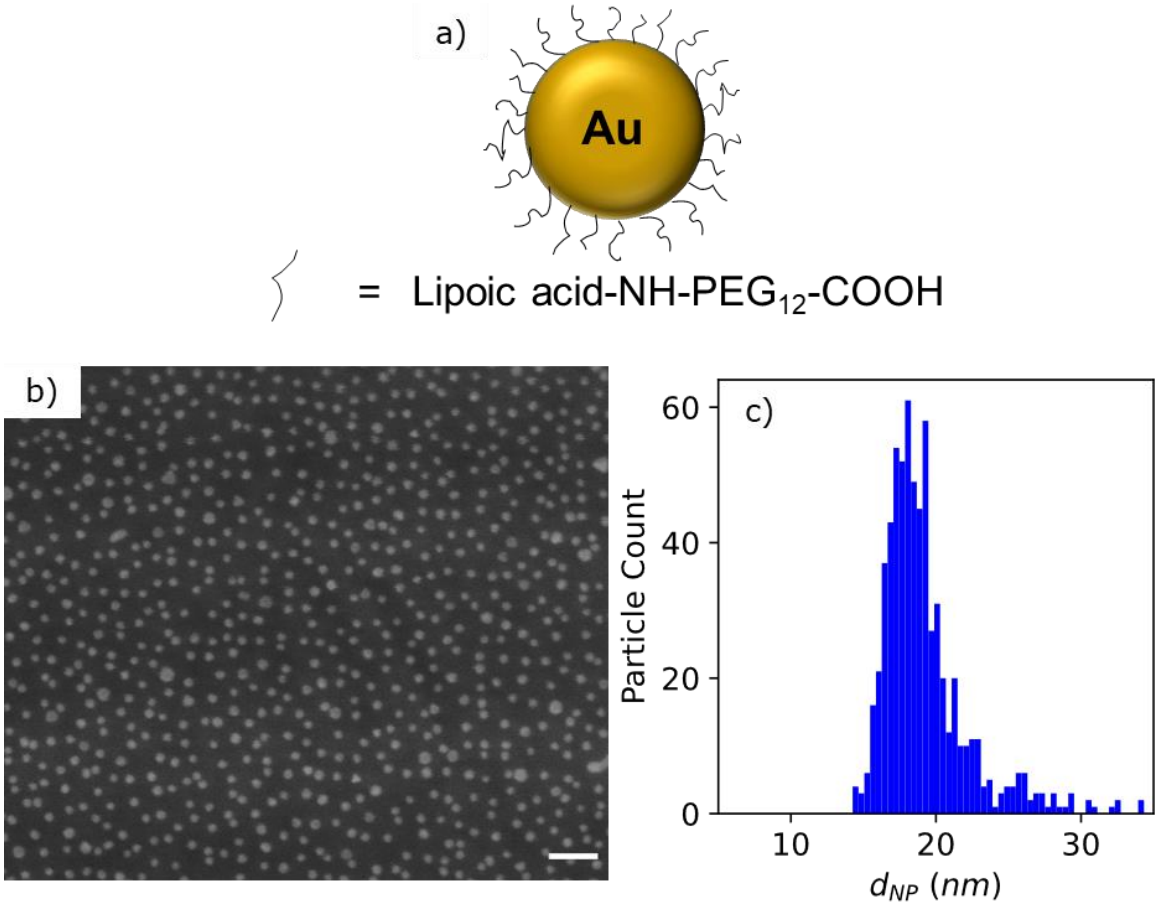


Figure S6. (a) AuNP particle and ligand chemistry. (b) SEM image of the same AuNPs imaged in the iSCAT experiment. Scale bar is 100 nm. (c) Particle size distribution histogram obtained by measuring the segmented region properties of particles in the SEM image (b).

Measuring Hydrodynamic Diameter with DLS ($d_{H,DLS}$)

For DLS measurements, 200 pM solutions of AuNPs in 1 mM Na_2PO_4H buffer at a pH of 7.0 were prepared and filtered. Malvern Zetasizer Nano S was used to measure the hydrodynamic diameter of particles. Temperature was maintained at 21.0º C and 5 independent

measurements of 70 seconds were collected with 30 second intervals between measurements. Number density distributions of particle size were determined using gold properties (n = 0.2, k = 3.320). Particle hydrodynamic radius was determined to be 20.1 ± 4.9 nm.

Section S5: Interferometric Scattering Microscopy

Interferometric Scattering Microscopy Instrumentation

An inverted iSCAT microscope configuration used for experiments was constructed following previously published literature[5,6] (See figure 3 in ref. S5 for schematic) and key details are described below. The collimated, coherent output of a fiber coupled laser (Coherent, OBIS 488-120 LS FP) at a wavelength of 488 nm forms the beam used for all imaging. The beam was passed through an orthogonal pair of acousto-optical deflectors (AODs, Gooch & Housego) and optically steered through a polarizing beam splitter (PBS) and a quarter-wave-plate (QWP) to the back focal plane of an oil immersion microscope objective (Nikon Apo TIRF, 1.49 NA, 100x). This resulted in a weakly focused 1-2 μm diameter beam rapidly scanning across the sample to generate an image. AODs were controlled by custom LabVIEW software developed in the Goldman Lab to sweep the set field of view within one exposure for each frame. Reflections from the sample-solution interface and backscattered light from the nanoparticles were collected by the objective. The reflected and scattered light was separated from the incident light by the PBS and QWP. Lenses and mirrors steered the reflected light to a CMOS detector (PhotonFocus AG MV1-D1024E-160). Images at the detector were collected using custom LabVIEW software provided from ref. S5 and modified to store the timestamp of video collection. All optical components sat on an active optical table to minimize vibrations. Lastly, the objective focus and stage position were controlled in three dimensions coarsely using stepper motors (OptoSigma SGSP-25ACTR-BO) and finely with a piezoelectric stage (PI USA, P-545.3C7). Stepper motors and the

piezoelectric stage were operated by a custom LabVIEW program previously developed by the Goldman Lab. The pixel size in our build was measured using a microscope stage micrometer to be 50 nm/pixel (~200x magnification).

Interferometric Scattering Microscopy Imaging

256 x 256 pixel images were collected resulting in a 12.8 μm x 12.8 μm field of view. The exposure time was set to 5.44 ms resulting in a frame rate of 184 fps including readout time. Laser intensity was set at the beginning of the experiment to achieve a maximum reflected intensity on the detector of 3000 (12-bit camera digitizer, 4096 intensity resolution). The focal plane was set using piezoelectric motors to maximize the sharpness (determined by eye during data collection) of background features in images. Data was collected in 1000 frame videos. Every 30 minutes, the focal plane was reset to account for drift. After the focal plane was reset, a static background image was collected by laterally translating the sample in a 2D Lissajous pattern using the piezoelectric motors and calculated as the pixelwise median of 100 images collected during that movement.[5] 12 videos (~ 66 seconds of experiment time) were collected using custom python scripts. This was repeated throughout the course of a 4 hour experiment. The ambient temperature of the room was recorded throughout the experiment and remained constant at 21 ± 1°C.

Section S6: Image Flattening

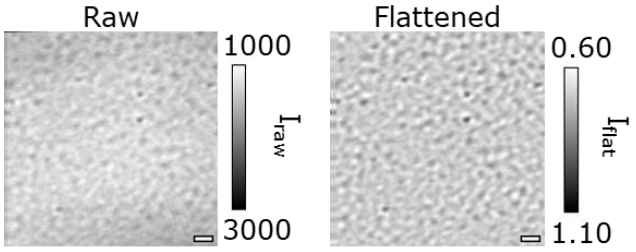


Figure S7. Raw and flattened iSCAT images of an AuNPs (19.2 nm) adsorbing onto an 85 nm LbL film (R_{rms} = 3.2 nm). (Left) A raw image of a LbL film on a glass coverslip in buffer solution. (Right) A flattened image produced by dividing the raw image by a static background image followed division by a median filtered image (kernel = 21). Note that in raw and flattened images, no particles can be seen. The rough LbL films scatter with stronger intensity (~0.40 relative to the background mean of 1) than the AuNPs (~0.04). All scale bars are 1 μm.

Section S7: Ratiometric Processing and Single Particle Tracking to Quantify Adsorption Event Time and Scattering Contrast

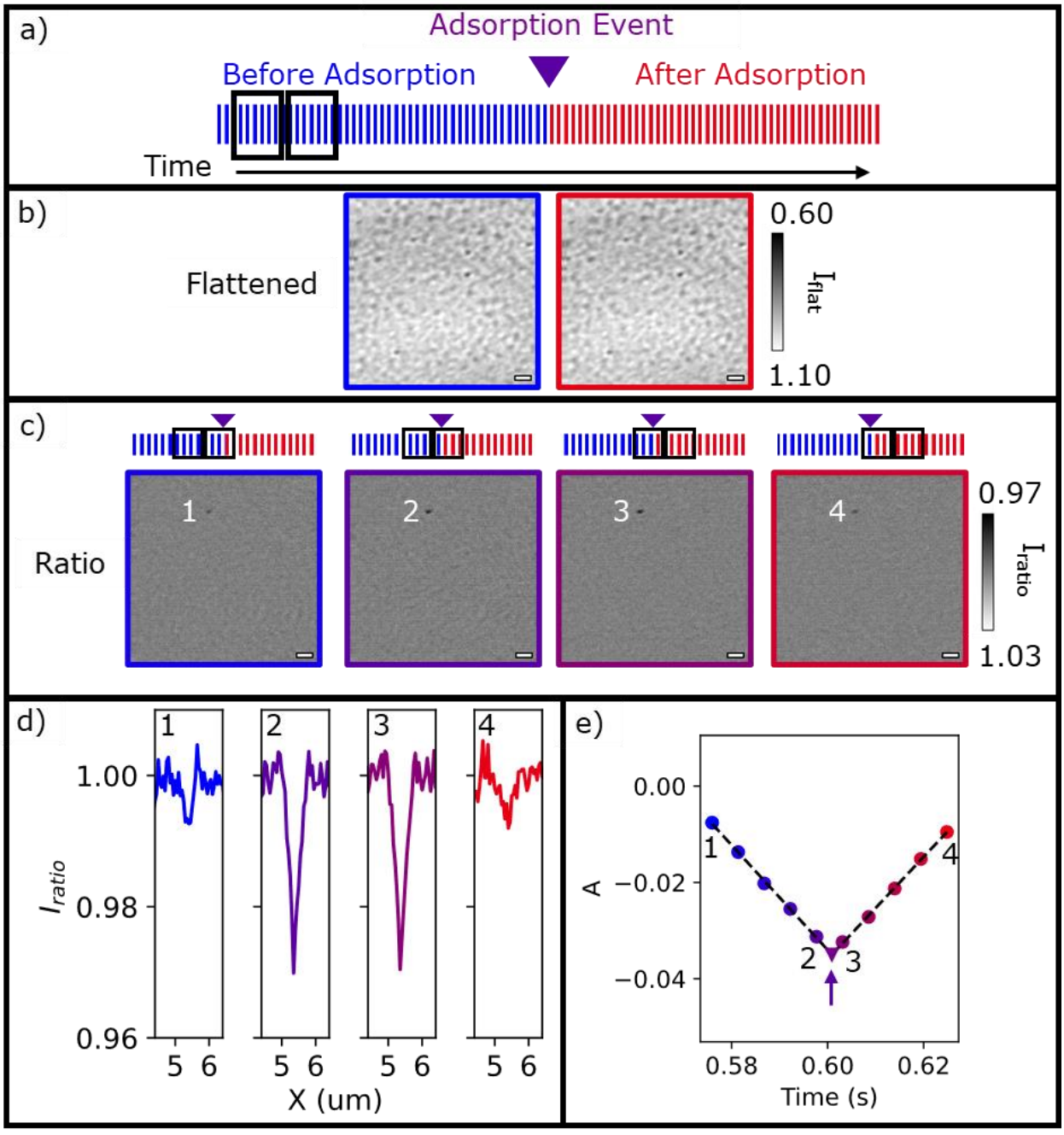


Figure S8. Ratiometric processing of a 19.2 nm AuNP adsorbing to an 85 nm LbL film (R_{rms} = 3.2 nm). (a) An illustration of ratiometric processing. The black boxes show the time-binning windows of size of N_{ratio} = 5. Blue vertical lines correspond to flattened frames prior to

the adsorption event (indicated by the purple triangle) and red vertical lines correspond to flattened frames after the event occurred. The black arrow shows increasing time. (b) Flattened iSCAT images before (blue border) and after (red border) the adsorption event. No particles are visible in the flattened images. (c) An illustration of the ratiometric processing approach. Above each image, the purple triangle marks the adsorption event. Flattened images in each box are averaged in the two boxes (left and right), normalized by their means, and the right average is divided by the left. As the midpoint between the boxes increases in time, the PSF from the adsorption event increases and decreases in contrast (darkness). (d) Corresponding cross-sections along the x direction through the y-centroid of the PSF for PSFs numbered in the ratiometric images. (e) Amplitudes (A) obtained from the 2D-Gaussian fits detected by the Haar or Mask R-CNN methods are linked into a trajectory. Numbers correspond to cross-sections in (d) and the PSF in (c). The dashed black line shows the fits of A versus time. The intersection of these lines (purple arrow) determines the scattering contrast ($|A_{event}|$) and binding time of the adsorption event. All scale bars are 1 μm.

Section S8: Mask R-CNN Training

After each epoch in the training process, weights were saved, the loss was calculated, and the mean average precision (AP; equations S2-S4**)** over all images in the training and validation sets was calculated.

$$AP = \frac{1}{11} \sum_{recall \in \{0,0.1,\ldots,1\}}^{recall} Precision(recall) \qquad \text{(S2)}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad \text{(S3)}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad \text{(S4)}$$

The precision and recall are analyzed at various classification prediction confidence thresholds. In iSCAT image analysis, a true positive was a particle detected and classified correctly as a dark or bright particle. A false positive was region of pixels pertaining to the background identified

incorrectly as being a positive (misclassifications errors, i.e., labeling a dark particle bright or vice versa, were not observed). The loss curve for the training process on a dataset containing representative backgrounds is shown in figure S9.

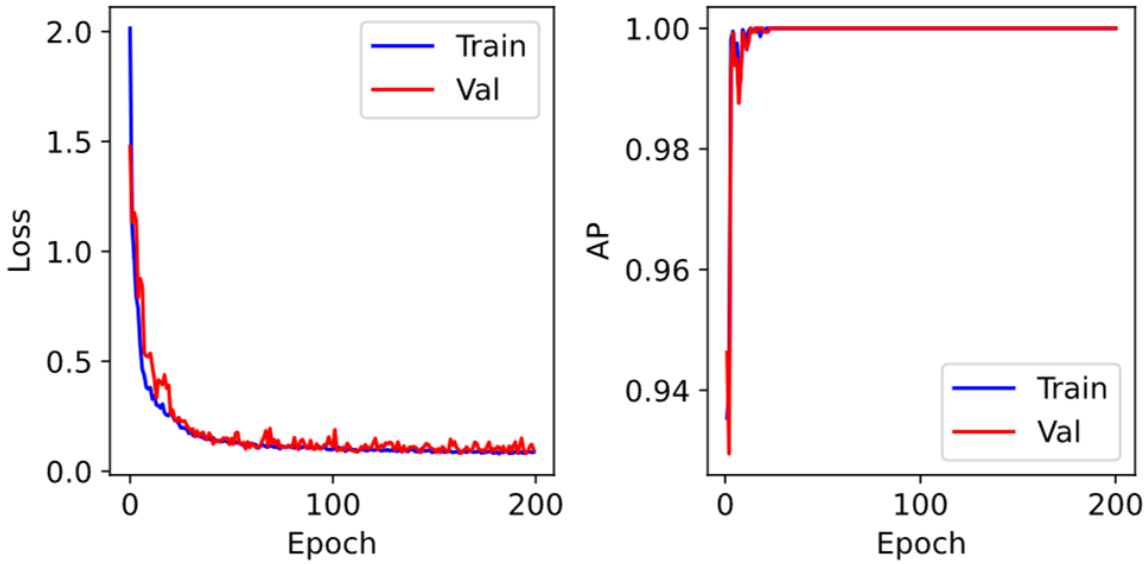


Figure S9. (left) Loss during training and validation evaluated using trained weights after each epoch. Monotonic decay of the loss function during training and validation indicated the network continued learning through the 200th epoch. The overlap of validation and training loss show no signs of overfitting. (right) Average precision throughout training and validation calculated after each epoch. Saturation occurs after ~50 epochs, but performance does not degrade after continued training.

The inputs and output of the inference of a trained mask R-CNN network on iSCAT dataset images are shown in figure S10.

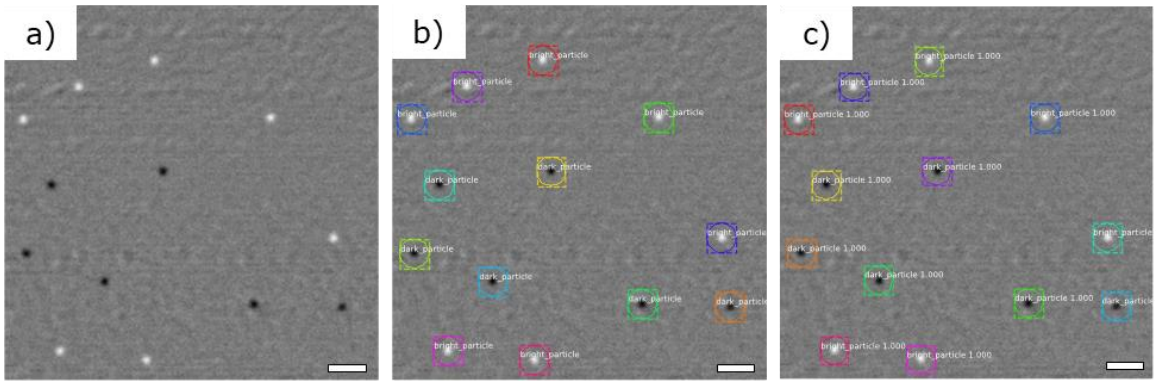


Figure S10. (a) A dataset image containing a ratiometric iSCAT background with synthetic particles. The image has been rescaled so that intensity varies from 0 to 1 and converted to an RGB image to match mask R-CNN input formatting. (b) The ground truth mask, bounding boxes, and classes for particles in image (a). (c) The output of predictions (mask, bounding box, class) of the trained mask R-CNN network when analyzing the image in (a). Numbers right of the predicted class indicate the confidence level of the classification.

Section S9: Effect of SNR_{min} on Particle Detection During Ratiometric Imaging

To illustrate the performance of each method in object detection and subsequent particle tracking, candidate detections with an SNR_{min} (defined in equation S5) less than a value of 1, 2, or 3 or are discarded prior to trajectory linking in ratiometric analysis.

$$SNR = \frac{|A|}{\sigma_{background}} \qquad \text{(S5)}$$

A is the amplitude from a 2D-Gaussian PSF fit of a detected particle, and $\sigma_{background}$ is the standard deviation of all background pixels in the image in which the proposed particle is detected. The sensitivity of the Haar and mask R-CNN methods in object detection to the setting of SNR_{min} is highlighted in figure S11.

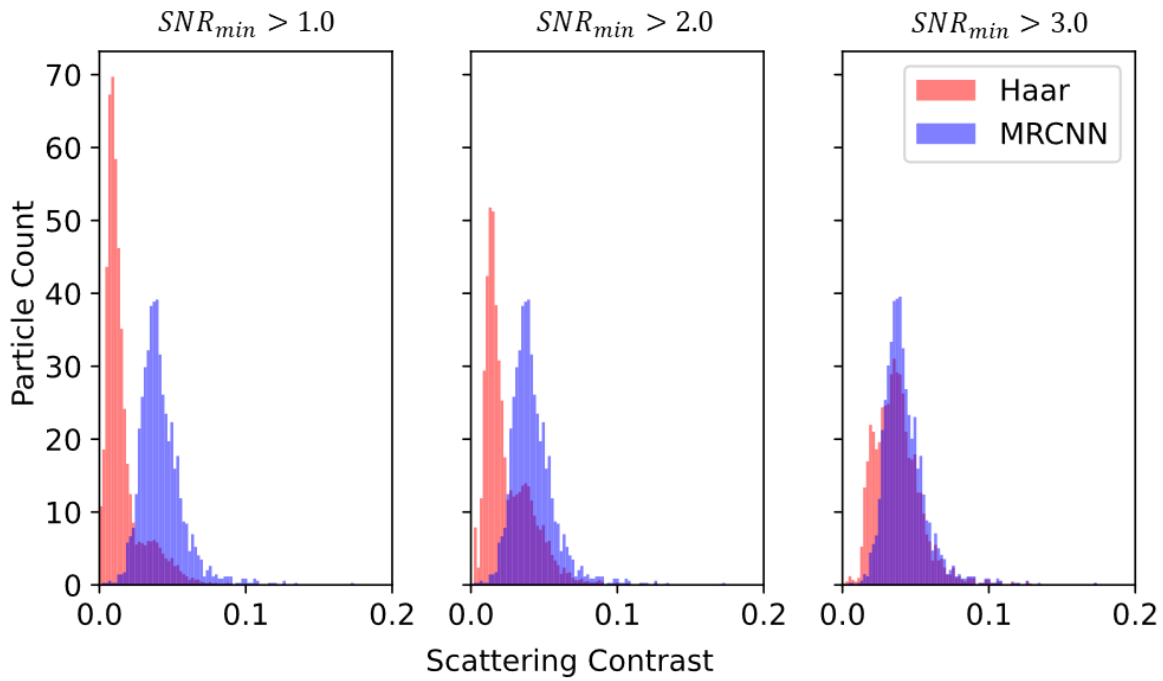


Figure S11. Histograms of ratiometric scattering contrast for varying values of SNR_{min}. A low contrast peak in Haar detections (pink) at all values of SNR_{min} is indicative of false positives with decreasing prominence as SNR_{min} increases. False positive detections also negatively impact linking of particle trajectories, resulting in a reduced number of correctly identified particle scattering contrasts in the central peak. Mask R-CNN detections (blue), showing a single peak, are invariant to this parameter highlighting its robust performance in particle detection at all SNR_{min} values.

Figure S11 illustrates two key additional concepts, First, the high density of false positive detections present in the Haar histograms have a negative impact on particle tracking linking algorithms. Incorrect linking of particles and background features into trajectories leads to failed fits during ratiometric analysis and an underestimation of true positives in the final analysis. Errors in particle tracking are also evident in more failed linear fits in ratiometric analysis as a function of frame or time (Haar = 1642 failed fits, mask R-CNN = 14 failed fits). Second, the Haar

histograms indicate the algorithm is very sensitive to the filtering parameter SNR_{min} while the mask R-CNN method is invariant to this parameter. This invariance emphasizes the ability of mask R-CNN to identity particles in a complex background at low signal to noise ratios.

Section S10: References

(1) Yoo, D.; Shiratori, S. S.; Rubner, M. F. Controlling Bilayer Composition and Surface Wettability of Sequentially Adsorbed Multilayers of Weak Polyelectrolytes. *Macromolecules* **1998**, *31* (13), 4309–4318. https://doi.org/10.1021/ma9800360.

(2) Lee, H. S.; Eckmann, D. M.; Lee, D.; Hickok, N. J.; Composto, R. J. Symmetric PH-Dependent Swelling and Antibacterial Properties of Chitosan Brushes. *Langmuir* **2011**, *27* (20), 12458–12465. https://doi.org/10.1021/la202616u.

(3) Stalder, A. F.; Kulik, G.; Sage, D.; Barbieri, L.; Hoffmann, P. A Snake-Based Approach to Accurate Determination of Both Contact Points and Contact Angles. *Colloids Surfaces A Physicochem. Eng. Asp.* **2006**, *286* (1–3), 92–103. https://doi.org/10.1016/j.colsurfa.2006.03.008.

(4) der Walt, S.; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2*, e453.

(5) Ortega Arroyo, J.; Cole, D.; Kukura, P. Interferometric Scattering Microscopy and Its Combination with Single-Molecule Fluorescence Imaging. *Nat. Protoc.* **2016**, *11* (4), 617–633. https://doi.org/10.1038/nprot.2016.022.

(6) Andrecka, J.; Takagi, Y.; Mickolajczyk, K. J.; Lippert, L. G.; Sellers, J. R.; Hancock, W. O.; Goldman, Y. E.; Kukura, P. *Interferometric Scattering Microscopy for the Study of Molecular Motors*, 1st ed.; Elsevier Inc., 2016; Vol. 581. https://doi.org/10.1016/bs.mie.2016.08.016.