Supporting Information

AutoDetect-mNP: an unsupervised machine learning algorithm for automated analysis of transmission electron microscope images of metal nanoparticles

Xingzhi Wang,[†,‡,¶] Jie Li,[†,‡,§] Hyun Dong Ha,[‡,¶] Jakob C. Dahl,[‡,¶] Justin C. Ondry,[‡,‖] Ivan Moreno-Hernandez,[‡] Teresa Head-Gordon,[*,‡,§,⊥,#] and A. Paul Alivisatos[*,‡,@,¶,‖]

†*These authors contributed equally*

‡*Department of Chemistry, University of California, Berkeley, California 94720, United States*

¶*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

§*Kenneth S. Pitzer Theory Center, University of California, Berkeley, California 94720, United States*

‖*Kavli Energy NanoScience Institute, Berkeley, California 94720, United States*

⊥*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

#*Departments of Bioengineering and Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States*

@*Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States*

E-mail: thg@berkeley.edu; paul.alivisatos@berkeley.edu

Supplementary Figures

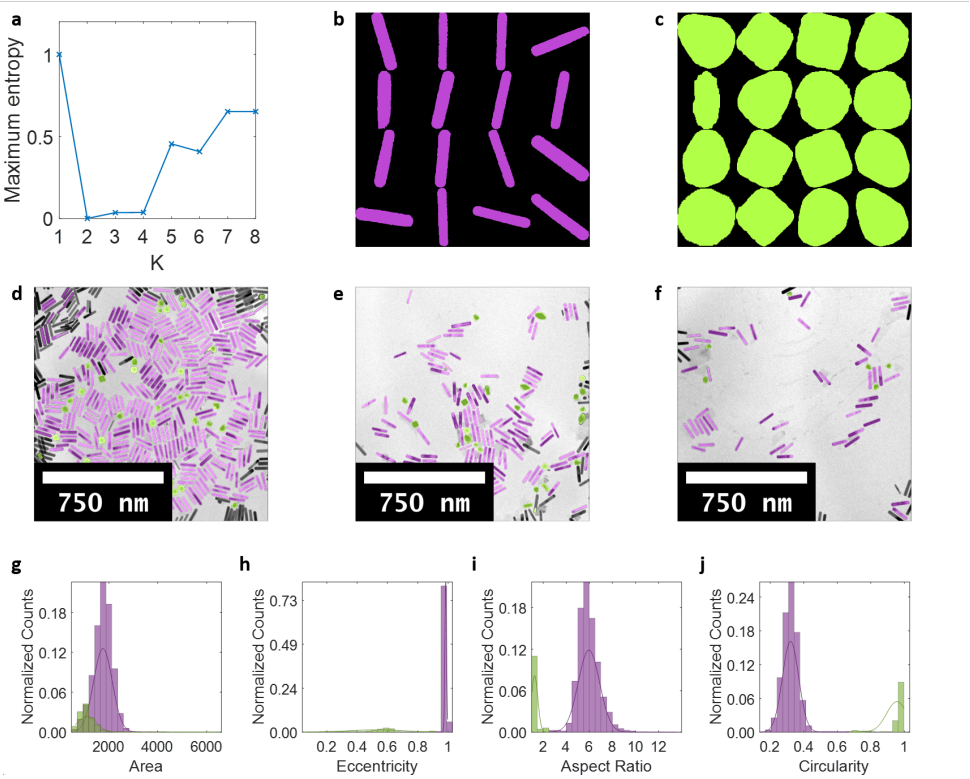

Figure S1: *Detection and classification of Au NPs of different morphologies in long rods.* (a) Maximum entropy as a function of number of classes in which K = 2 was found to be the optimal number of classes. (b-c) Montages of sample particle shapes in each class. (d-f) Classification results denoted by colors overlaid onto original TEM images of Au nanorods (green: spheroids, purple: long rods). (g-j) Four features used for classification and Gaussian distributions for each class, with classification results denoted by colors. Counts normalized by total number of particles.

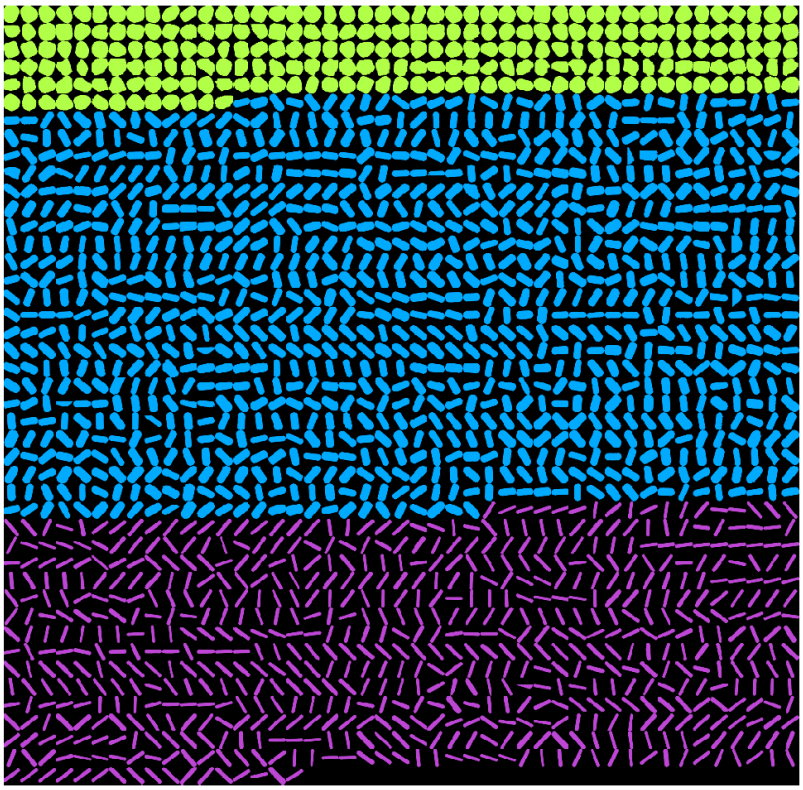

Figure S2: *Montage of all particles detected in the mixture sample.* Classification result is denoted by color (purple: long rods, green: spheroids, blue: short rods).

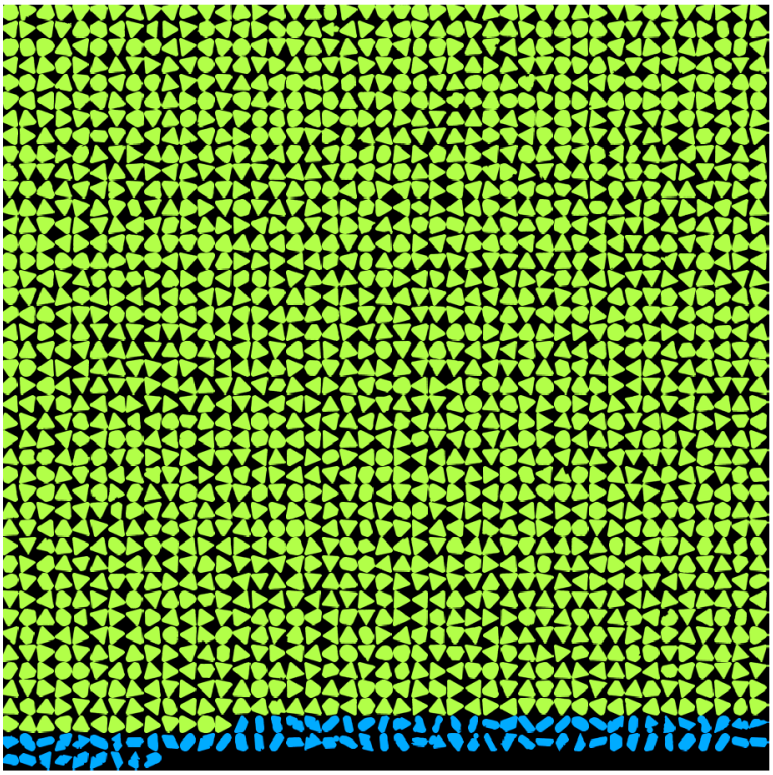

Figure S3: *Montage of all particles detected in the triangular prisms sample.* Classification result is denoted by color (green: triangular particles, blue: rod-shaped impurities).

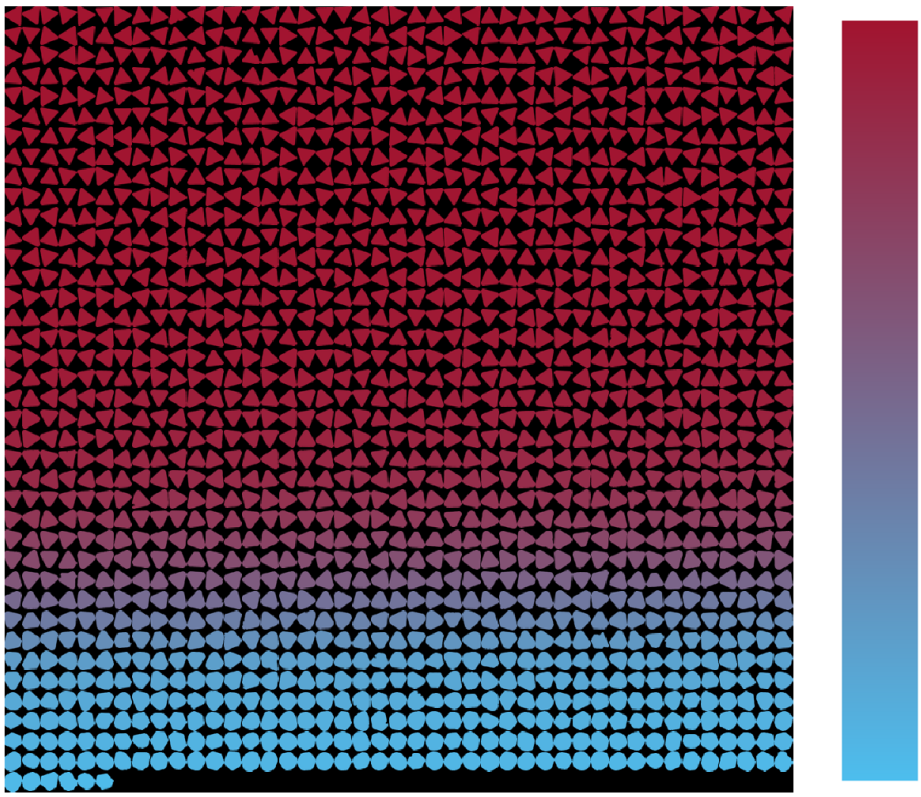

Figure S4: *Transition between triangular shapes and hexagonal shapes in the triangular prisms sample.* Red color represents the particle's likelihood of belonging to the class of the triangular shapes. Blue color represents the particle's likelihood of belong to the class of the hexagonal shapes.

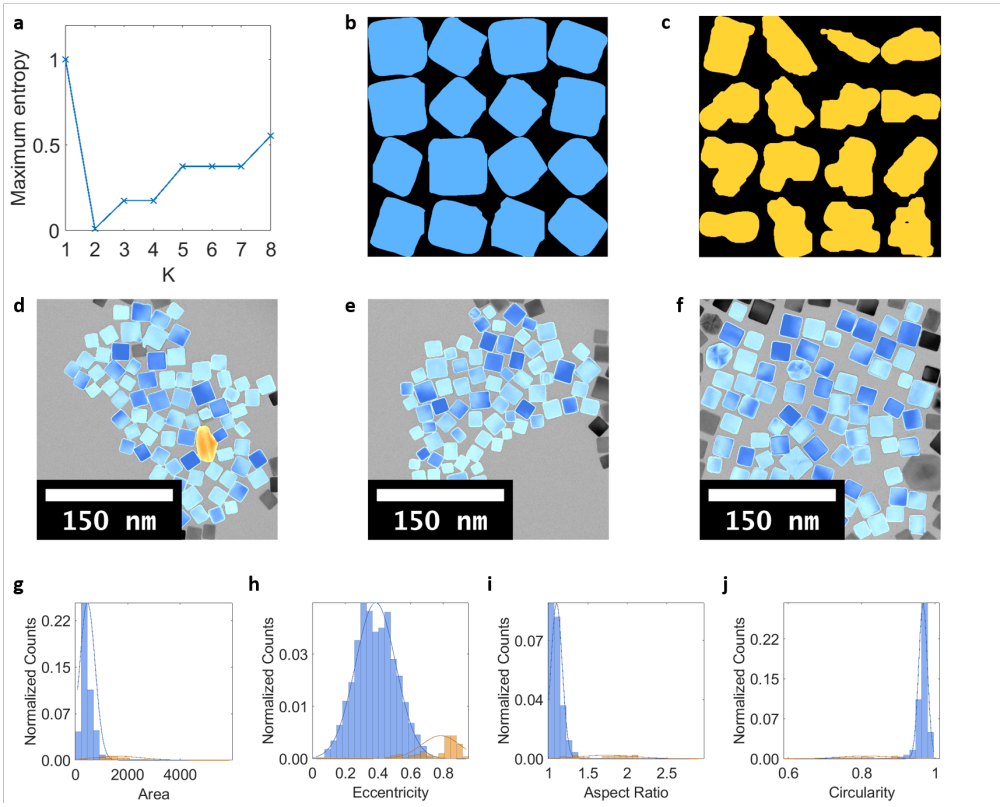

Figure S5: *Detection and classification of Pd nanocubes of different morphologies.* (a) Maximum entropy as a function of number of classes in which K = 2 was found to be the optimal number of classes. (b-c) Montages of sample particle shapes in each class. (d-f) Classification results denoted by colors overlaid onto original TEM images of Pd nanocubes (blue: cubes, orange: irregularly shaped impurities). (g-j) Four features used for classification and Gaussian distributions for each class, with classification results denoted by colors. Counts normalized by total number of particles.

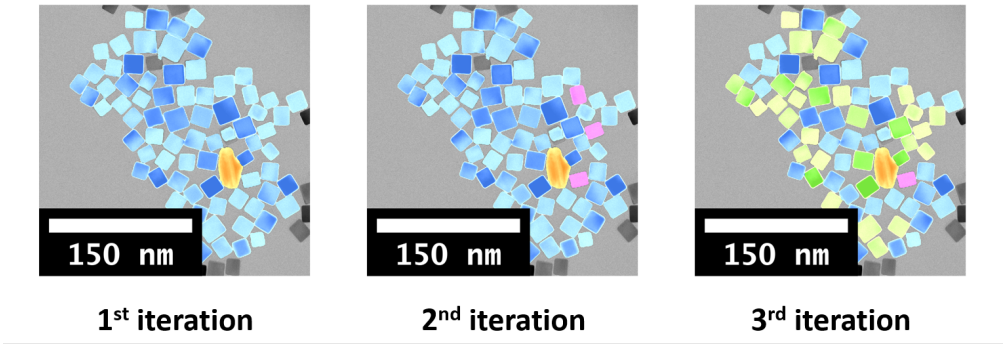

Figure S6: Classification results of Pd nanocubes with different numbers of iterations.

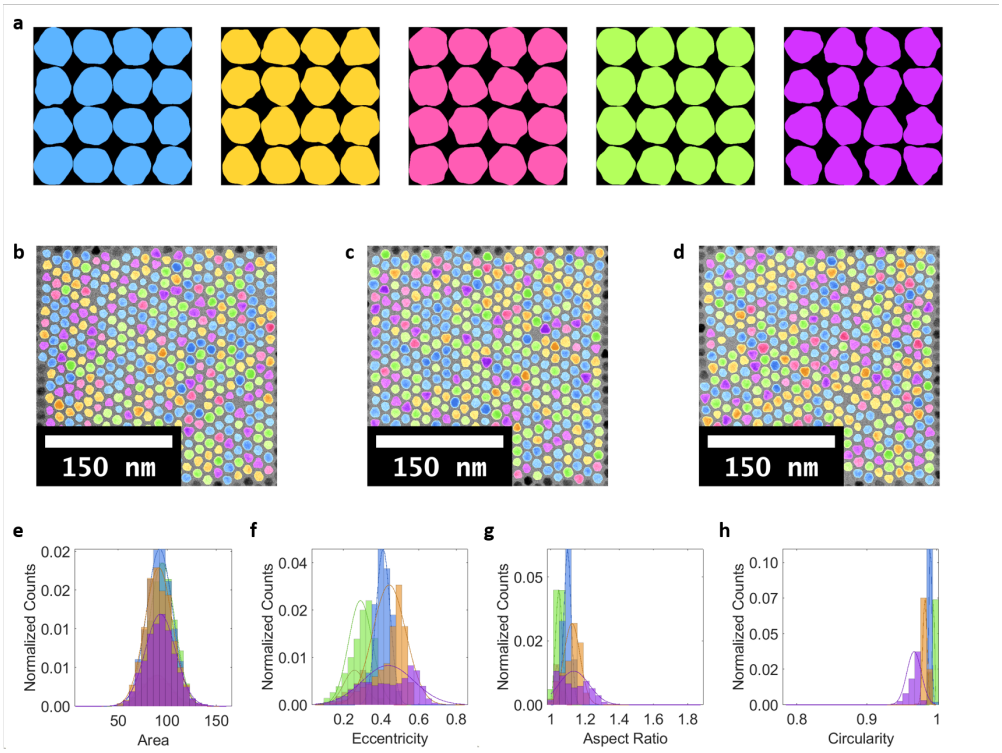

Figure S7: *Detection and classification of CdSe/CdS QDs of different morphologies.* (a) Montages of sample particle shapes in each class. (b-d) Classification results denoted by colors overlaid onto original TEM images of CdSe QDs. (e-h) Four features used for classification and Gaussian distributions for each class, with classification results denoted by colors. Counts normalized by total number of particles.

Supplementary Tables

Table S1: Relative populations (mean ± standard deviation) of Au NPs of different morphologies in the three samples: mixture, short rods, and long rods. For each sample, all images were randomly split into three subsets, and populations of NPs with different morphologies were counted for each subset, and means and standard deviations of the populations were calculated from the three subsets. Based on the results shown above, it can be calculated that the mixture sample is consisted of 45% of long rods sample and 55% of short rods sample.

	Long rod	Short rod	Sphere
Mixture sample	0.39 ± 0.05	0.52 ± 0.05	0.09 ± 0.005
Long rod sample	-	0.95 ± 0.006	0.05 ± 0.005
Short rod sample	0.85 ± 0.02	-	0.15 ± 0.03

Table S2: Relative polulations (mean ± standard deviation) of Au NPs of different morphologies claculated by AutoDetect-mNP and human labellers. Data were collected from three human labellers working independently and the algorithm analyzing the same set of 20 images of the mixture sample.

	Long rod	Short rod	Sphere
Human labeller	0.43 ± 0.009	0.506 ± 0.008	0.091 ± 0.002
AutoDetect-mNP	0.38	0.52	0.10

Table S3: Cross entropy of each classes with respect to the bulk for the mixture sample and triangular prisms sample.

Triangular prisms	Cross entropy	Mixture	Cross entropy
Asymmetrically truncated triangles	1.89	Spheroids	4.93
Symmetrically truncated triangles	2.26	Short rods	2.36
Triangles	2.76	Long rods	7.84
Rod-shaped impurities	1.47		