# 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,[†,‡] Teresa Head-Gordon,[*,†,¶,§,‖] and A. Paul Alivisatos[*,†,⊥,‡,#]

†*Department of Chemistry, University of California, Berkeley, CA, USA*

‡*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

¶*Kenneth S. Pitzer Theory Center, University of California, Berkeley, CA, USA*

§*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

‖*Departments of Bioengineering and Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA*

⊥*Department of Materials Science and Engineering, University of California, Berkeley, CA, USA*

#*Kavli Energy NanoScience Institute, Berkeley, CA, USA*

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

**Abstract**

The synthesis quality of artificial inorganic nanocrystals is most often assessed by transmission electron microscopy (TEM) for which new high-throughput advances have dramatically increased both the quantity and information richness of metal nanoparticle (mNP) characterization. Existing automated data analysis algorithms of TEM

mNP images generally adopt a supervised approach, requiring a significant effort in human preparation of labelled data that reduces objectivity, efficiency, and generalizability. We have developed an unsupervised algorithm AutoDetect-mNP for automated analysis of TEM images that objectively extracts morphological information of convex mNPs from TEM images based on their shape attributes, requiring little to no human input in the process. The performance of AutoDetect-mNP is tested on two datasets of bright field TEM images of Au nanoparticles with different shapes, demonstrating that the algorithm is quantitatively reliable, and can thus serve as a generalizable measure of the morphology distributions of any mNP synthesis. The AutoDetect-mNP algorithm will aid in future developments of high-throughput characterization of newly synthesized mNPs, and the future advent of time-resolved TEM studies that can investigate reaction mechanisms of mNP synthesis and reactivity.

# Introduction

The shape attributes of a metal nanoparticle (mNP) determine many of its physical, chemical, and functional properties, including plasmonic behavior,[1–4] catalytic efficiency,[5–8] and biological activity.[9,10] Significant research effort has focused on developing methods for the shape-controlled synthesis of mNPs, and an accurate and efficient method for their characterization has been indispensable for the success achieved by these studies.[11–18] To date, transmission electron microscopy (TEM) remains the most reliable and widely used method for characterizing the morphology of NPs, for which rapid advancement in automated high-throughput electron microscopy has drastically increased both the acquisition rate and the quality of TEM data.[19–23] Increased efficiency in TEM data acquisition now enables the scale of NP shape characterization to increase from tens or hundreds of particles to orders of magnitude more, extracting information at the level of more informative *statistical distributions*. The information from TEM data at such a scale far exceeds the capability of a human analyst, and hence the development of automated methods for TEM image analysis

is imperative.

An ideal algorithm for the automated analysis of TEM images should be able to perform two tasks: particle detection, which includes identification and segmentation of particles of interest, and information extraction, which generally involves the characterization of the shapes of the particles based on their attributes such as aspect ratios and other attributes. While a variety of robust algorithms have been developed to perform automated particle detection,[21,24–29] these algorithms were not designed to perform statistical information extraction and data analysis that identify morphological features of the NPs. When information extraction has been automated, it has generally analyzed all the detected particles without differentiation or classification of their shape and shape attributes.[21,24,25] Such an approach may be acceptable when analyzing homogeneous samples with the aim of quantifying simple metrics such as average particle sizes. However, the assumption of homogeneity is subjective, and far from universally applicable. By taking the assumption of homogeneity, other potentially important information may be missed, including the presence of unexpected side products, or the relative populations of different competing products, which may be lost without detailed automated classification of particle shapes. Existing algorithms capable of classifying the particles based on their shapes have thus far adopted supervised approaches, requiring the user to pre-define expected categories or shapes of particles.[26,27] Such supervised methods reduce the generalizability of these algorithms, often limiting their application to the particular classes of particles on which they were trained. To our best knowledge, there has not been an unsupervised algorithm capable of analyzing TEM images without significant human input or presumptions.

Here, we develop an unsupervised algorithm for the analysis of TEM images and the classification of particle shapes for mNPs, requiring minimum human input in the process (Figure 1). We demonstrate the quantitative reliability of the AutoDetect-mNP algorithm and its potential to serve as an unbiased, general method for the characterization of the shape distributions of nanoparticles using two datasets of gold nanoparticles (AuNPs) with

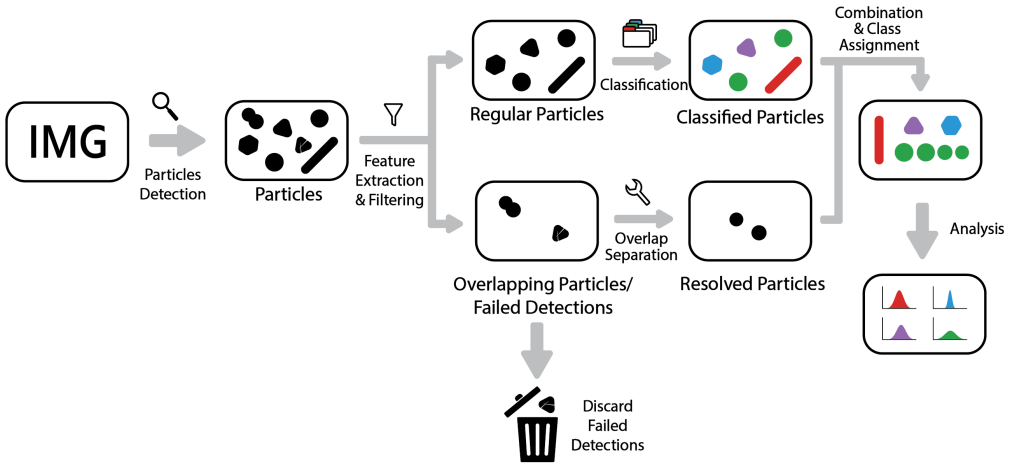


Figure 1: *Scheme of the AutoDetect-mNP algorithm.* The algorithm can be divided into four parts: particle detection, feature extraction, filtering and resolution of irregularly shaped particles, classification of particle shapes.

different shapes comprised of short and long rods and triangular prisms. AuNPs have been studied extensively by TEM, and protocols for their synthesis with controlled sizes and shapes are well established,[12,15–17,30–34] making them an ideal candidate for demonstration of the strengths of the unsupervised algorithm. Furthermore, while shape-controlled synthesis of AuNPs have been extensively reported in literature, a generalizable quantitative method for the characterization of the shape distribution of these particles on a statistical scale beyond manual analysis of TEM images has yet to be established.[12,15–17,30–33] AutoDetect-mNP is capable of processing thousands of particles from over a hundred TEM images in a matter of minutes, extracting morphological descriptors for convex Au nanoparticles and classifying the particles without the need of human intervention, thus greatly increasing the generalizability and efficiency of the process. In the future, AutoDetect-mNP can serve as a general and unbiased metric for reporting shape and shape attribute distributions of mNPs using TEM or other image acquisition strategies, and will play an important role in the future development of automated platforms for high-throughput mNP synthesis and time-resolved TEM characterization of mNP reactivity.

# Results and discussion

## AutoDetect-mNP algorithm

Our goal is to develop an unsupervised algorithm capable of achieving four tasks in an automated manner for mNP particle classification. The components of our algorithm are individual mNP detection, feature extraction of particle shape attributes, the resolution and filtering out of overlapping particles, and finally shape classification and the analysis of their distributions that are interpretable from their features.

Particle detection of mNP requires the segmentation of the foreground from the background of the TEM images, and thus the isolation of individual particle from the segmented image. Since TEM images of Au NPs typically show sharp contrast between the foreground and background, well-developed algorithms reported in previous literature generally shows satisfactory performance.[26] In this study, K-means image segmentation has proven sufficient for this step. Six shape descriptors are used to distinguish the mNPs that are segmented from the images: solidity, convexity, area, eccentricity, aspect ratio, and circularity, .[26,35–38]

The first two descriptors are important for cleaning the dataset for mNP detection by measuring the degree of convexity of the particles. Since a TEM image can only capture the 2D projections of 3D particles, overlapping particles will appear as the union of their projections. Additionally, particle detection can sometimes fail to render the complete edges of certain particles, or identify contamination on TEM sample grids as particles. Fortunately, overlapping and incorrectly detected particles often have non-convex shapes, allowing them to be easily differentiated from correctly detected mNPs, which typically have convex shapes. In this work, we apply a filter based on solidity and convexity to separate non-convex particles from correctly detected convex particles. A particle with solidity $< 0.9$ or convexity $< 0.95$ is considered non-convex. Non-convex particles are analyzed and resolved by the ultimate erosion of convex shapes (UECS) algorithm reported by Park et al.[39] Although we only quantify convex mNPs in this work, which represent a significant proportion of widely studied

mNPs, future research efforts can use a similar algorithm and different descriptors to extend the application of the algorithm to non-convex NPs.

Particles that pass the solidity and convexity filter are then classified based on the 4 geometrical shape descriptors. The classification process consists of two unsupervised steps: K-means clustering followed by naive Bayes classification. This general approach of unsupervised classification based on geometric descriptors ensures that it can be naturally extended to any convex NP systems. Data points were divided into K classes, assuming features in each class follow a normal distribution. Each data point was then assigned to the class that maximizes the joint probability of all the features belonging to that class. In a standard implementation of K-means clustering and naive Bayes classifier, the number of classes, K, needs to be pre-defined by the user. To increase the generalizability of our algorithm, we use the $P_{max}$ metric to automatically determine the optimal K for each individual dataset. For each dataset, matrix $\boldsymbol{P}$(K) is defined as a function of the number of classes, K, such that

$$P_{ij} = \frac{\mathrm{E}[p_j(x_i)]}{\mathrm{E}[p_i(x_i)]}, i, j \leq K$$

, where $P_{ij}$ is the element on the $i$-th row and $j$-th column of $\boldsymbol{P}$(K), $p_i(x)$ is the likelihood of data point $x$ belonging to class $i$, and $x_i$ denotes data points assigned to class $i$. Maximum entropy $P_{max}$(K) is then defined as the maximum of the off-diagonal elements of $\boldsymbol{P}$(K). The optimal K is then selected as the value of K that minimizes $P_{max}$(K).

## Au nanorods

To confirm the reliability of the AutoDect-mNP algorithm, we first analyzed TEM images of two pure Au nanorod samples, each with a different aspect ratio of ∼5 (long rods) and ∼2 (short rods), the synthesis of which was previously reported by Ye et al.[32] The algorithm yields statistical results on the pure nanorod samples that consistently show nanorods with their known aspect ratios with narrow size and shape distributions as shown in Figure 2

(short rods) and Figure S1 (long rods). However, we also found a non-negligible amount of spheroidal impurities present in both samples. Due to the relatively low concentrations of these nanorod samples, it is unlikely that these spheroidal particles are Au nanorods standing on their tips. We note that the presence of spheroidal impurities during the synthesis of Au nanorods has been previously observed,[11,17,32] and the amount of impurities are typically estimated by optical spectroscopy or visual inspection of a few electron microscopic images,[12,16,17,30–32] which is both time consuming and prone to human biases. AutoDetect-mNP determines the impurities of the Au nanorods to be 15% in long rods and 5% in short rods. This shows that the algorithm can potentially serve as a quantitative method for the objective determination of impurity content of a mNP synthesis.

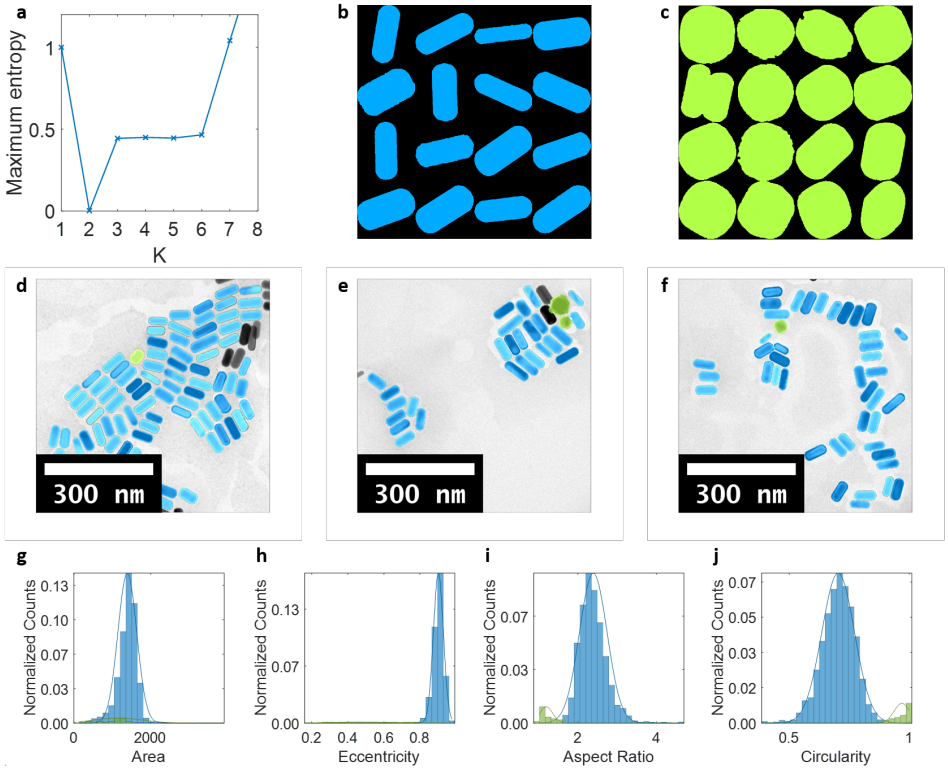


Figure 2: *Detection and classification of Au NPs of different morphologies in short rods.* (a) Maximum entropy as a function of number of classes in which K = 2 was found to be the optimal number of classes. (b-c) Montages of sample particle shapes in each class. (d-f) Classification results denoted by colors overlaid onto original TEM images of Au nanorods (green: spheroids, blue: short 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.

Next we examined the accuracy of the algorithm on a sample with a known 1:1 mixture composition of two Au nanorods with different aspect ratios, determined from spectroscopic measurement of their colloidal concentrations using computed extinction coefficients.[32] The performance of AutoDetect-mNP on the nanorod mixture is shown in Figure 3 in which it automatically identifies three classes among all the recognized particles as long rods, short rods, and spheroids. Visual inspection of the distributions of the shape descriptors (Figure 3h-k) shows that, similar to the cases of the pure rods (Figure 2g-j, Figure S1g-j), the most prominent distinctions differentiating the three classes are seen in aspect ratio and circularity. To test the quantitative performance of the algorithm, the numbers of detected particles in each class were counted for the mixture sample and compared to the two pure nanorod samples (Table S1). We find that the composition of the nanorod mixture sample is calculated to be 45% long rods and 55% short rods, which shows satisfactory accuracy compared to the spectroscopically determined 1:1 ratio between the two samples.

The performance of AutoDetect-mNP was also compared to that of three human experts who were asked to independently classify the particles into three predefined classes: long rods, short rods, and spheroids found in a set of 20 images randomly selected from the nanorod mixture dataset. The total number of particles recognized by the algorithm (860) was very close to the number recognized by human experts (915±25). The distinct numbers of long rods and short rods identified by the algorithm were found to be very close to the average of human labelling results, with small deviations (Table S2). While the difference between the numbers of spheroids identified by human labellers and the algorithm was larger (12.8%), this is likely due to the fact that the spheroids have a relatively small absolute population (88±15 according to human labelling). As such, the algorithm is capable of achieving a similar degree of accuracy as human experts, while being significantly more efficient.

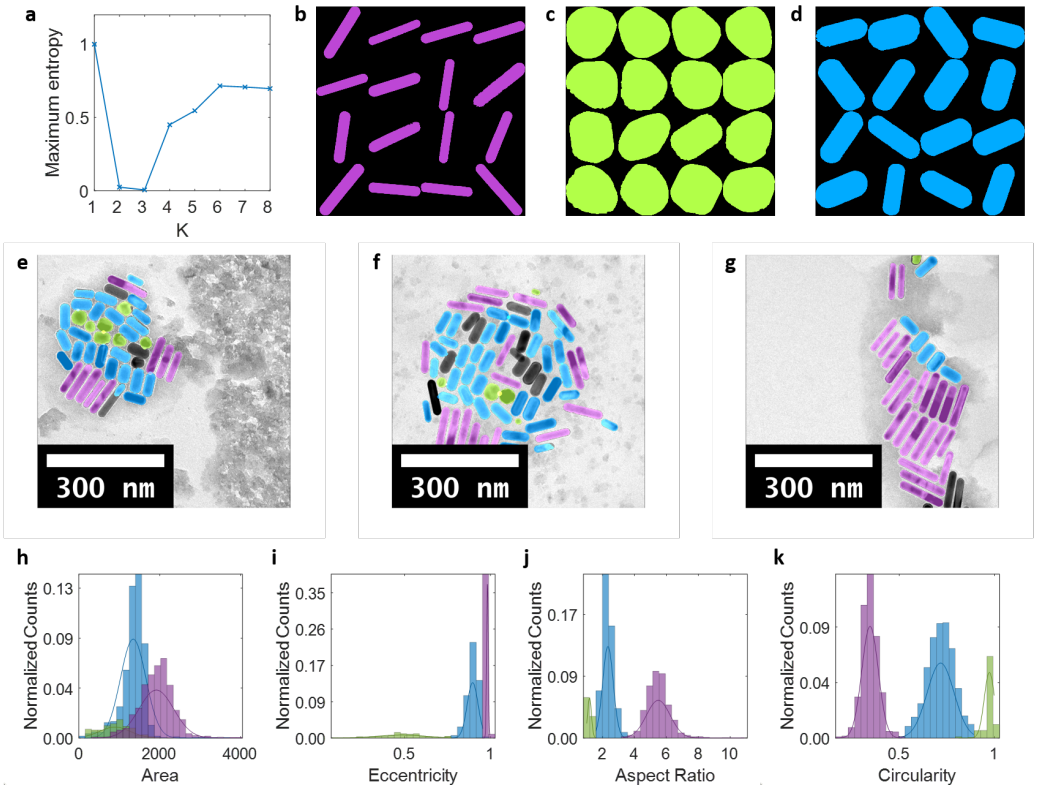


Figure 3: *Classification results of a 1:1 mixture of Au nanorods with different aspect raios* (a) Maximum entropy as a function of number of classes in which K = 3 was found to be the optimal number of classes. (b-d) Montages of sample particle shapes in each class. (e-g) Sample TEM images of the mixture with classification results denoted by color (purple: long rods, green: spheroids, blue: short rods). (h-k) Four features used for classification and Gaussian distributions for each class, with classification results denoted by colors. Counts normalized by total number of particles.

## Au triangular prisms

After validating the performance of the algorithm on an artificially constructed data set, we demonstrate the potential of AutoDetect-mNP for characterizing and quantifying mNPs for a new synthesis of Au triangular prisms as reported by Jones et al.[33] We chose Au triangular prisms since they show a relatively high variation in shapes from the synthesis.[11,15,33,40] From the 114 TEM images collected for the product, the AutoDetect-mNP algorithm identified two classes from the detected particles: a class of triangular particles (Figure 4b) and a class of rod-shaped impurities (Figure 4c). By visually inspecting the rod-shape impurities, it can be seen that a majority of the rod-shaped mNPs are in fact real impurities, with a much smaller population of artifacts originating from incorrectly detected particles (Figure 4c). We would like to point out that the relative population of the incorrectly detected particles is low enough that their presence would not influence the classification results on a statistical scale.

While the two classes identified by the algorithm show clear distinctions, further inspection of the class of triangular particles (Figure 4b) reveals that there is significant intraclass variations within the triangular particles, as evident from the non-Gaussian distributions of the features (Figure 4g-j).To quantify the distinctiveness of the two classes, we evaluate the average cross entropy from information theory, where the average cross entropy between two distributions $p$ and $q$ is defined as $H(p,q) = \frac{1}{N}\sum_i^N p_i \log(q_i)$, where $N$ is the number of elements in distribution $p$, and $p_i$ and $q_i$ represent the likelihood of the i-th element of distribution $p$ predicted by distribution $p$ and $q$, respectively. $H(p,q)$ quantifies the distinctiveness between distribution $p$ and distribution $q$: the larger $H(p,q)$, the more distinct $p$ is from $q$. Using this measure, the two classes identified from the triangular prism sample are found to be much less distinct than the nanorod case. This can be seen quantitatively by comparing the average cross entropy with respect to the bulk distribution (Table S3). We hypothesize that the distinctions between the triangular particles and rod-shaped impurities were so large that the possible presence of intraclass variations among the triangular particles

were weighted to a lesser degree and thus not identified by the algorithm.

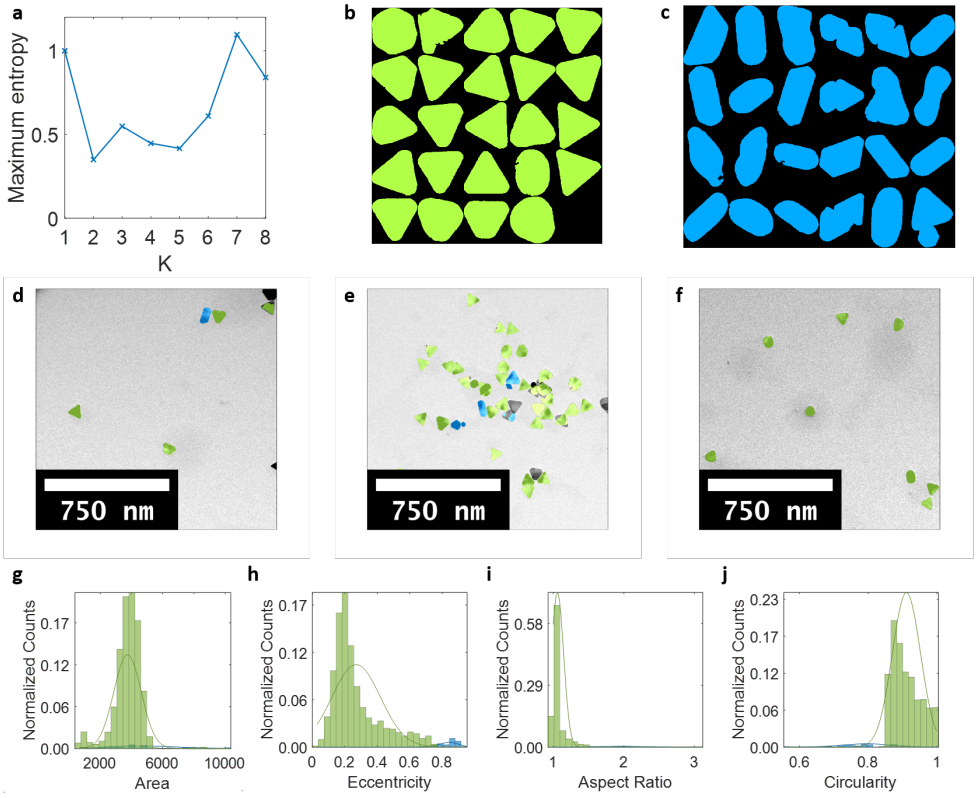


Figure 4: *Detection and classification of Au NPs of different morphologies in a sample of triangular prisms.* (a) Maximum entropy as a function of number of classes in which K = 2 was found to be the optimal number of classes. (b-c) Montages of sample particle shapes in each class. (d-f) Classification results denoted by colors overlaid onto original TEM images of Au nanorods (green: triangular particles, blue: rod-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.

To validate this idea, we applied the automatic classification of AutoDetect-mNP on the triangular particles alone in an attempt to further characterize the variations among these particles (Figure 5). The classification results showed that the triangular particles can indeed be further classified into three sub-classes: triangles, symmetrically truncated triangles, and asymmetrically truncated triangles (Figure 5). Compared to the 2-class scenario (triangles and rod-shaped impurities) in the first iteration of classification, classifying the particles into four classes (triangles, symmetrically truncated triangles, asymmetrically truncated triangles, and rod-shaped impurities) in the second iteration significantly improved log-likelihood

from 6.58 to 14.89, indicating that the second iteration of classification offered a more detailed and more reasonable classification of the particles in the dataset. Upon applying more iterations of classification, the log-likelihood did not improve further, suggesting there is no more significant intraclass variation to capture and that the algorithm has converged. Applying further iterations of classification on the class of rod-shaped impurities also did not improve the log-likelihood, suggesting that the class has little intraclass variation. In this case, the averaged cross entropy between subclasses are now sufficiently large such that these subclasses are quantifiably distinct from each other. Interestingly, inspection of the distributions of the shape descriptors for the three subclasses originated from the triangular shaped class (Figure 5) shows that none of the four shape descriptors can serve as a good differentiator for the three sub-classes alone. However, by taking into account of all four descriptors the algorithm was able to make a clear distinction between the three sub-classes. This demonstrates the importance of using high-dimensional features for the classification of shapes.

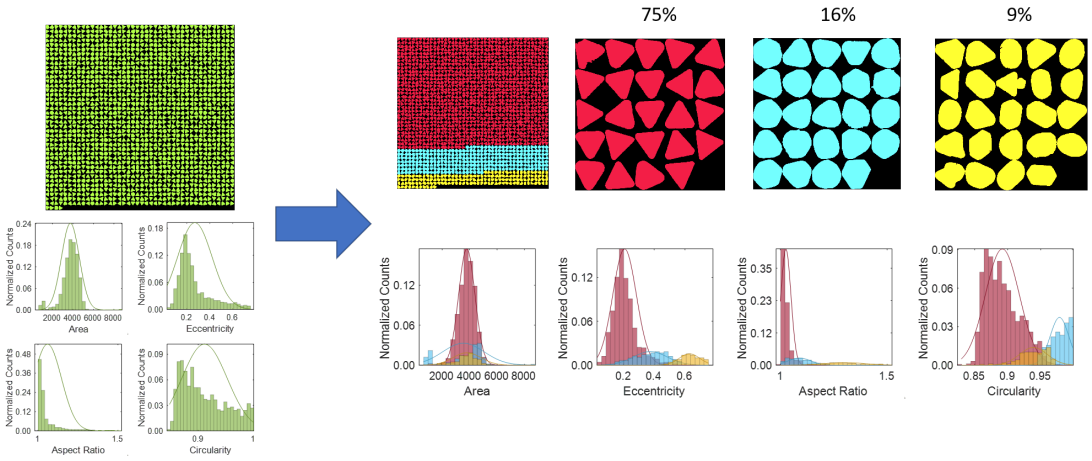


Figure 5: Further classification of the Au triangular prisms class. Distributions and relative population of particles in each class (red: pure triangles, cyan : symmetrically truncated triangles, yellow : asymmetrically truncated triangles)

The AutoDetect-mNP refined classification of truncated triangular particles in TEM images of Au triangular prisms is supported by previous studies characterizing these particles at both bulk[15,33,40–42] and single particle[14,26,43] levels. However Figure 5 also emphasizes

that the AutoDetect-mNP algorithm can provide the statistical distributions of each species and their descriptors. These information can potentially be used to interpret kinetically controlled formation mechanisms and relative speed of growth of different facets during the formation of Au nanoparticles by, for example, helping to illuminate the shape transition from three-fold symmetrical triangular particles to six-fold symmetrical hexagonal particles (Figure S4). In future studies, a kinetic growth model, potentially similar to the one previously reported by Handwerk et al.,[44] could be established based on the statistical distributions of particle shapes shown herein, which can potentially help to validate or improve existing models of nanoprism growth such as the twinned facets model proposed by Jaggannathan et al.[15,45]

# Conclusions

In this work, we developed an unsupervised machine learning algorithm, AutoDetect-mNP, that classifies Au NPs in TEM images based on extracted morphological information without human intervention. AutoDetect-mNP automatically decides the optimal number of classes for clustering based on intra- and inter-class feature similarities, and provides more relevant statistical distributions of the features for each class generated by the algorithm. The use of classes as a way of characterizing the shape distributions of nanoparticles can also be quantified with a cross-entropy measure post defacto to distinguish it from a more featureless continuum.

By comparing the analysis results on the composition of Au nanorod mixtures with different aspect ratios with human labelling, we demonstrated that our method is quantitatively reliable while being much more efficient, and potentially less biased than human labelling. Applying the algorithm to the characterization of shape distributions of synthesized Au triangular prisms, the algorithm identified rod-shaped impurities from the triangular main products. Furthermore, running the algorithm iteratively on identified classes can organize

the shapes of the particles hierarchically. It is hoped that the quantitative shape distributions provided by the algorithm can potentially help to provide insights into the mechanism of the formation of the nanoparticles.

As for now, the AutoDetect-mNP algorithm has only been applied to Au NPs. However, the geometric features utilized in the algorithm can be easily applied to other NP systems as well, and it is straightforward to add new features to expand the application domain of the algorithm. We can imagine future studies using AutoDetect-mNP to answer scientific questions such as optimizing reaction conditions for a high-throughput experimental setup, modelling reaction mechanism with real-time analysis of particle shapes captured by TEM videos, or studying the self-assembly of NPs with different shapes.

# Methods

## Chemicals and Materials

Hexadecyltrimethylammonium bromide (CTAB, >98.0%) and sodium oleate (NaOL, >97.0%) were purchased from TCI America. Hydrogen tetrachloroaurate trihydrate ($HAuCl_4 \cdot 3H_2O$, ≥99.9%), L-ascorbic acid (BioXtra, ≥99.0%), silver nitrate ($AgNO_3$, ≥99.0%), sodium borohydride ($NaBH_4$, 99.99%), sodium iodide (NaI, ≥99.5%), sodium hydroxide (NaOH, ≥97.0%), sodium citrate tribasic dihydrate (≥99.0%) and hydrochloric acid (36.5%-38.0% wt%) were obtained from Sigma Aldrich (USA). NaBH4 powder was stored in an argon glovebox. $HAuCl_4 \cdot 3H_2O$, L-ascorbic acid, and $AgNO_3$ were stored in a vacuum desiccator at room temperature. Deionized water (DI-water, Milipore, Milford, MA, USA) was used for all aqueous solution. All the glassware was thoroughly cleaned using freshly prepared aqua regia (3:1 volume ratio of HCl and $HNO_3$, respectively) followed by fully rinsing with copious amounts of DI-water. All chemicals were of reagent grade and used without further purification.

## Synthesis of gold nanorods

Two types of AuNRs were synthesized by a facile seed-mediated growth involving a binary surfactant mixture. The seed solution was prepared as follows: 10 mL of 0.1 M CTAB solution was mixed with 100 $\mu$L of 25 mM $HAuCl_4$ in a 20 mL scintillation vial under vigorous stirring. 600 $\mu$L of ice cooled 10 mM NaBH4 was rapidly injected into the Au-CTAB solution and stirred for 2 min. Upon the addition of $NaBH_4$, the color of the seed solution turned yellow-brownish. Afterward, the seed solution was left undisturbed at 28 °C for 30 min prior to use in the following step.

The growth solution was obtained by first mixing 3.6 g of CTAB and 0.4936 g of NaOL in 196 mL of DI-water in a 500 mL Erlenmeyer flask. The solution was heated with occasional agitation until all the CTAB was dissolved. The mixture was allowed to cool down to 30 °C and 4 mM $AgNO_3$ referred to Table 1 was then added under stir at 700 rpm for 15 min. Afterward, 4 mL of 25 mM $HAuCl_4$ was added to the mixture and kept undisturbed at 28 °C for 90 min. The yellowish color of growth solution turned to colorless. A certain amount of HCl (Table 1) was added to the solution and the mixture was stirred at 400 rpm for 15 min. Finally, 500 $\mu$L of 0.064 M ascorbic acid was injected into the growth solution, and the mixture was vigorously stirred at 1200 rpm for 30 sec. 80 $\mu$L of the seed solution was then injected, and the solution was stirred for 30 sec before left undisturbed at 28 °C for 12 hr to complete the growth process. 40 mL of the final products were isolated by centrifugation at 8,000 rpm for 15 min followed by careful removal of the supernatant. 10 mL of DI-water was added to the pellet and the mixture was sonicated briefly to disperse the pellet.

Table 1: Amount of reagents used for the synthesis of Au nanorods

| Sample | 4 mM $AgNO_3$ (mL) | HCl (mL) |
|---|---|---|
| Long rods | 1.45 | 0.84 |
| Short rods | 9.6 | 1.2 |

## Synthesis of gold triangular prisms

Homogeneous gold triangular prisms were also synthesized by seed-mediated method. The citrate ligand based seed solution was prepared as follows: 500 $\mu$L of 10 mM sodium citrate solution was mixed with 250 $\mu$L of 10 mM $HAuCl_4$ and 18.95 mL of DI-water in a 20 mL scintillation vial under vigorous stirring. 300 $\mu$L of ice cooled 10 mM $NaBH_4$ was rapidly injected into the Au-citrate solution and stirred for 1 min. Upon the addition of $NaBH_4$, the color of the seed solution turned yellow-brownish. Afterward, the seed solution was stirred at 40-45 °C for 15 min and left undisturbed at room temperature prior to use in the following step.

For the growing process, the UV-vis spectra of the seed solution were taken using 1 cm quartz cuvette to determine the concentration of the seed. The extinction coefficient of the seed solution is $9.696\times10^6$ $M^{-1}cm^{-1}$ at its wavelength of maximum optical density, 504 nm. In order to synthesize the triangular prism with edge length in 80 nm, the final concentration of the seed in the growth solution should be 97.7 pM. Prior to preparing the growth solution, the stock mixture of 0.05 M CTAB and 50 $\mu$M NaI was prepared at room temperature. The solution was heated with occasional agitation until all the CTAB was dissolved. The mixture was allowed to cool down to 30 °C. The growth solution was prepared by mixing 9 mL of previously made CTAB/NaI stock mixture, 250 $\mu$L of 10 mM $HAuCl_4$ solution, 50 $\mu$L of 100 mM NaOH solution, and 50 $\mu$L of 100 mM ascorbic acid solution. Finally, a certain amount of seed solution was added under vigorous stirring. The nanoparticle solution was then heated around 40 °C for 30 min and was cooled down to room temperature.

For the purification process of Au nanoprisms, 0.6 mL of the growing solution was mixed with 0.4 mL of 1 M NaCl solution into 1.5 mL Eppendorf tubes. The mixture was left undisturbed for 4 hours and centrifuged twice at 1,000 rpm for 15 sec. The supernatant was carefully removed, and 0.6 mL of DI-water was added to the pellet and was sonicated briefly to disperse the pellet.

## Preparation of TEM samples

To prepare the stock sample for TEM analysis, 50 $\mu$L aliquot of the nanoparticle solution was added to 1 mL of DI-water. The sample was centrifuged at 5,500 rpm for 8 min, the supernatant was carefully removed, and the isolated product was resuspended in 50 $\mu$L of DI-water. We added 200 $\mu$L of DI-water to the sample to dilute by a factor of 5 to prevent overlapping particles on the grid. 5 $\mu$L of this solution was pipetted onto a standard carbon TEM grid (Electron microscopy Sciences, CF-400-Cu). The TEM grid with the sample was fully dried in a vacuum desiccator at room temperature before TEM imaging.

In order to calculate the concentration of each nanorod, we referred to the theoretical extinction coefficient of the AuNR reported by Park et al.[46] Each extinction coefficient of long and short AuNR was extracted to $1.4 \times 10^{10}$ L mol$^{-1}$ cm$^{-1}$ at 1059 nm for long rods, and $1.1 \times 10^{10}$ L mol$^{-1}$ cm$^{-1}$ at 698 nm for short rods. The optical density of the solution was collected using a Shimadzu UV-3600 UV-VIS spectrophotometer with 1 nm resolution. The path length of the cuvette was 1 cm. To make 50:50 ratio of short and long AuNRs, the concentration of each sample was fixed to 90 pM.

## TEM imaging

Images of Au NPs were acquired using a FEI Tecnai T20 transmission electron microscope equipped with a Gatan RIO16IS camera and a $LaB_6$ filament. All images were recorded under 200 kV accelerating voltage. Drift correction feature of the imaging software was enabled during the acquisition of all images. For each sample, images were taken from a series of evenly spaced locations on the sample grid, typically resulting in 50-100 images of the sample containing $>$ 1000 recognizable NPs. We observed that magnification and camera exposure time can significantly impact the performance of particle detection. Therefore, it is potentially helpful to optimize these imaging parameters before beginning to collect large amounts of data for the algorithm. For the data showed in this work, long rods were imaged at 43kx magnification with 1 s exposure time; short rods were imaged at 19kx magnification

with 1 s exposure time; mixtures were imaged at 43kx magnification with 1 s exposure time; triangular prisms were imaged at 26kx magnification with 1.5 s exposure time.

## AutoDetect-mNP algorithm

Unprocessed TEM images in dm4 format were loaded into Matlab by the *ReadDMFile* function available from MathWorks file exchange. Particle detection was performed by Matlab's built-in function *imsegkmeans*. *Imsegkmeans* segments the image by performing a K-means clustering on the pixel intensities of each image. Shape descriptors for each particles were calculated by the *regionprops* function in Matlab. The definition for each shape descriptors is as follows: Area is defined as the total area, in $nm^2$, of the region encompassed by the edge of the particle; eccentricity is defined as the eccentricity of the smallest ellipse that encapsulates the particle; aspect ratio is defined as the ratio between the major and minor axis lengths of the particle; circularity is defined as the reciprocal of the 1st Hu's moment invariant of the particle;[37,38] solidity and convexity, are defined as the ratio of the areas and perimeters, respectively, between the particle and the smallest convex polygon that encapsulates the particle. Hu's moment invariants of the particles were calculated by codes available from MathWorks file exchange.

UECS was implemented using Matlab codes adapted from those published by Park et al.[39] The UECS algorithm attempts to resolve convex components from non-convex shapes by iteratively eroding the shapes until convex shapes are obtained. If convex shapes are not obtained after 90% of the area of the original particle has been eroded, the particle will be discarded.

K-means clustering and naive Bayes classifier used for unsupervised classification are well established clustering algorithms, and are implemented in Matlab. An example demonstrating using $P_{max}$(K) to decide the optimal number of classes is shown in Figure 6. In Figure 6, circularity of the Au nanorods was used as an example to demonstrate the selection of optimal K. In this case, K = 3 is selected as the optimal K for this dataset, which agrees

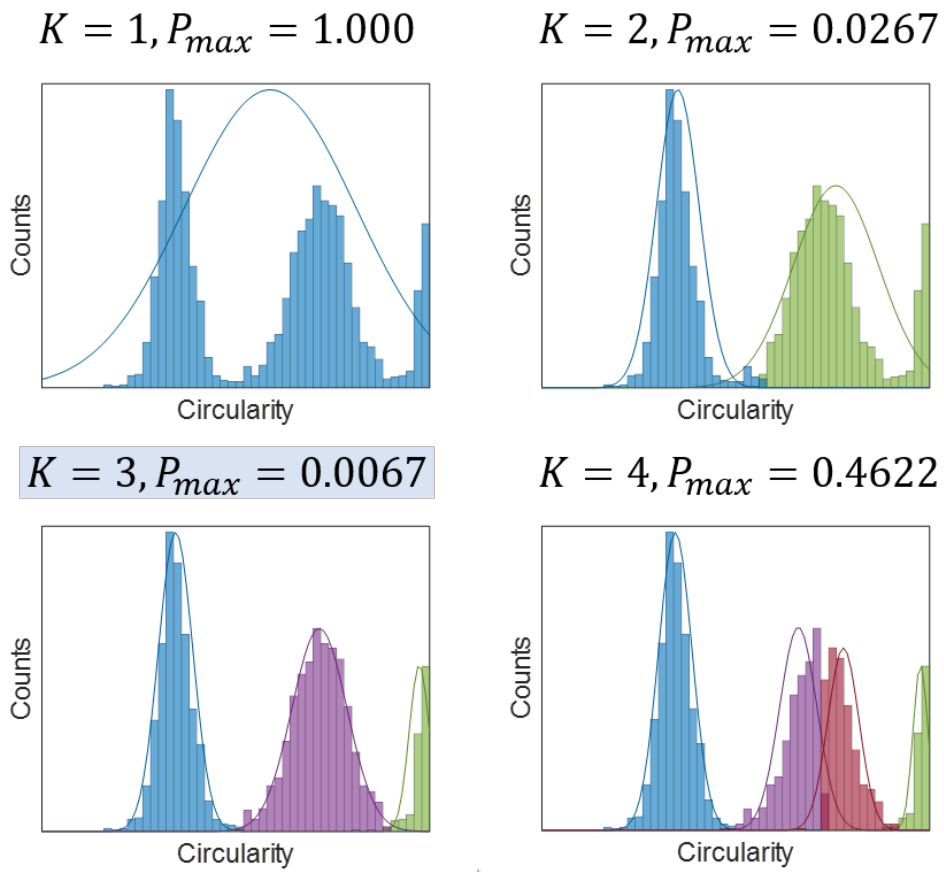


Figure 6: Unsupervised clustering of extracted features and selection of optimal number of clusters (K). Using circularity of Au nanorods as an example.

with the observation that the dataset is roughly consisted of three normal distributions. Of note, particles resolved by UECS were excluded during the automated classification process to avoid skewing of classification results by potential artefacts generated by UECS. Instead, these particles were assigned classes after the normal particles are clustered into different classes.

## Acknowledgement

X.W., H.D.H., J.C.D. and A.P.A. thank the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, for their support for the nanoparticle synthesis and TEM imaging part of this work under Contract No. DE-AC02-05-CH11231 within the Data Science for Discovery in Chemical and Materials Sciences Program (KCD252). J.L. and T.H-.G. thank the National Institutes of Health for their support for the computational and algorithm development part of this work under Grant No 5U01GM121667. We also thank Priyanka Raghavan, Vishnu Dharmaraj, Megan Lim, Adam Rodgers, and Hao Lyu for early work as part of Chem195/295: Machine Learning for the Chemical Sciences taught by T.H-G.. Many users have provided feedback on the class, which is reflected in all of the different demonstrations shown in this document.

## Author Contributions

X.W., J.L., A.P.A. and T.H-G. conceived the scientific direction; X.W., J.L., T.H-G., H.D.H., and J.C.D. wrote the manuscript; X.W. and J.L. provided calculations and created the Figures. All authors contributed insights and discussed and edited the manuscript.

## Declaration of Interests

The authors declare no competing interests.

## Supporting Information Available

Additional data and visualization of results, including Figure S1-4 and Table S1-4, are available in the Supporting Information.

## References


(1) Noguez, C. Surface plasmons on metal nanoparticles: The influence of shape and physical environment. *Journal of Physical Chemistry C* **2007**, *111*, 3606–3619.

(2) Kelly, K. L.; Coronado, E.; Zhao, L. L.; Schatz, G. C. The optical properties of metal nanoparticles: The influence of size, shape, and dielectric environment. *Journal of Physical Chemistry B* **2003**, *107*, 668–677.

(3) Mock, J. J.; Barbic, M.; Smith, D. R.; Schultz, D. A.; Schultz, S. Shape effects in plasmon resonance of individual colloidal silver nanoparticles. *J. Chem. Phys.* **2002**, *116*, 6755.

(4) Orendorff, C. J.; Sau, T. K.; Murphy, C. J. Shape-Dependent Plasmon-Resonant Gold Nanoparticles. *Small* **2006**, *2*, 636–639.

(5) Narayanan, R.; El-Sayed, M. A. Catalysis with transition metal nanoparticles in colloidal solution: Nanoparticle shape dependence and stability. *Journal of Physical Chemistry B* **2005**, *109*, 12663–12676.

(6) Narayanan, R.; El-Sayed, M. A. Shape-dependent catalytic activity of platinum nanoparticles in colloidal solution. *Nano Letters* **2004**, *4*, 1343–1348.

(7) Lykaki, M.; Pachatouridou, E.; Carabineiro, S. A.; Iliopoulou, E.; Andriopoulou, C.; Kallithrakas-Kontos, N.; Boghosian, S.; Konsolakis, M. Ceria nanoparticles shape effects on the structural defects and surface chemistry: Implications in CO oxidation by Cu/CeO2 catalysts. *Applied Catalysis B: Environmental* **2018**, *230*, 18–28.

(8) King, M. E.; Personick, M. L. Defects by design: Synthesis of palladium nanoparticles with extended twin defects and corrugated surfaces. *Nanoscale* **2017**, *9*, 17914–17921.

(9) Xie, X.; Liao, J.; Shao, X.; Li, Q.; Lin, Y. The Effect of shape on Cellular Uptake of Gold Nanoparticles in the forms of Stars, Rods, and Triangles. *Scientific Reports* **2017**, *7*, 1–9.

(10) Chithrani, B. D.; Ghazani, A. A.; Chan, W. C. Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. *Nano Letters* **2006**, *6*, 662–668.

(11) Ha, T. H.; Koo, H. J.; Chung, B. H. Shape-controlled syntheses of gold nanoprisms and nanorods influenced by specific adsorption of halide ions. *Journal of Physical Chemistry C* **2007**, *111*, 1123–1130.

(12) Lohse, S. E.; Murphy, C. J. The quest for shape control: A history of gold nanorod synthesis. *Chemistry of Materials* **2013**, *25*, 1250–1261.

(13) Grzelczak, M.; Pérez-Juste, J.; Mulvaney, P.; Liz-Marzán, L. M. Shape control in gold nanoparticle synthesis. *Chemical Society Reviews* **2008**, *37*, 1783–1791.

(14) Wang, S. T.; Lin, Y.; Nielsen, M. H.; Song, C. Y.; Thomas, M. R.; Spicer, C. D.; Kröger, R.; Ercius, P.; Aloni, S.; Stevens, M. M. Shape-controlled synthesis and in situ characterisation of anisotropic Au nanomaterials using liquid cell transmission electron microscopy. *Nanoscale* **2019**, *11*, 16801–16809.

(15) Millstone, J. E.; Hurst, S. J.; Métraux, G. S.; Cutler, J. I.; Mirkin, C. A. Colloidal gold and silver triangular nanoprisms. *Small* **2009**, *5*, 646–664.

(16) Requejo, K. I.; Liopo, A. V.; Zubarev, E. R. Gold Nanorod Synthesis with Small Thiolated Molecules. *Langmuir* **2020**, *36*, 3758–3769.

(17) Vigderman, L.; Khanal, B. P.; Zubarev, E. R. Functional gold nanorods: Synthesis, self-assembly, and sensing applications. *Advanced Materials* **2012**, *24*, 4811–4841.

(18) Sau, T. K.; Murphy, C. J. Room temperature, high-yield synthesis of multiple shapes of gold nanoparticles in aqueous solution. *Journal of the American Chemical Society* **2004**, *126*, 8648–8649.

(19) Mulligan, S. K.; Speir, J. A.; Razinkov, I.; Cheng, A.; Crum, J.; Jain, T.; Duggan, E.; Liu, E.; Nolan, J. P.; Carragher, B.; Potter, C. S. Multiplexed TEM Specimen Preparation and Analysis of Plasmonic Nanoparticles. *Microscopy and Microanalysis* **2015**, *21*, 1017–1025.

(20) Tan, Y. Z.; Cheng, A.; Potter, C. S.; Carragher, B. Automated data collection in single particle electron microscopy. *Microscopy (Oxford, England)* **2016**, *65*, 43–56.

(21) Schorb, M.; Haberbosch, I.; Hagen, W. J.; Schwab, Y.; Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nature Methods* **2019**, *16*, 471–477.

(22) Kim, B. H. et al. Critical differences in 3D atomic structure of individual ligand-protected nanocrystals in solution. *Science* **2020**, *368*, 60–67.

(23) House, S. D.; Chen, Y.; Jin, R.; Yang, J. C. High-throughput, semi-automated quantitative STEM mass measurement of supported metal nanoparticles using a conventional TEM/STEM. *Ultramicroscopy* **2017**, *182*, 145–155.

(24) Sosa, J. M.; Huber, D. E.; Welk, B.; Fraser, H. L. Development and application of MIPAR$^{TM}$: a novel software package for two- and three-dimensional microstructural characterization. *Integrating Materials and Manufacturing Innovation* **2014**, *3*, 123–140.

(25) Cervera Gontard, L.; Ozkaya, D.; Dunin-Borkowski, R. E. A simple algorithm for measuring particle size distributions on an uneven background from TEM images. *Ultramicroscopy* **2011**, *111*, 101–106.

(26) Laramy, C. R.; Brown, K. A.; O'Brien, M. N.; Mirkin, C. A. High-Throughput, Algorithmic Determination of Nanoparticle Structure from Electron Microscopy Images. *ACS Nano* **2015**, *9*, 12488–12495.

(27) Boselli, L.; Lopez, H.; Zhang, W.; Cai, Q.; Giannone, V. A.; Li, J.; Moura, A.; de Araujo, J. M.; Cookman, J.; Castagnola, V.; Yan, Y.; Dawson, K. A. Classification and biological identity of complex nano shapes. *Communications Materials* **2020**, *1*, 1–12.

(28) Groschner, C. K.; Choi, C.; Scott, M. C. Methodologies for Successful Segmentation of HRTEM Images via Neural Network. *arXiv* **2020**,

(29) Qian, Y.; Huang, J. Z.; Li, X.; Ding, Y. Robust nanoparticles detection from noisy background by fusing complementary image information. *IEEE Transactions on Image Processing* **2016**, *25*, 5713–5726.

(30) Murphy, C. J.; Thompson, L. B.; Chernak, D. J.; Yang, J. A.; Sivapalan, S. T.; Boulos, S. P.; Huang, J.; Alkilany, A. M.; Sisco, P. N. Gold nanorod crystal growth: From seed-mediated synthesis to nanoscale sculpting. *Current Opinion in Colloid and Interface Science* **2011**, *16*, 128–134.

(31) Zhao, P.; Li, N.; Astruc, D. State of the art in gold nanoparticle synthesis. *Coordination Chemistry Reviews* **2013**, *257*, 638–665.

(32) Ye, X.; Zheng, C.; Chen, J.; Gao, Y.; Murray, C. B. Using binary surfactant mixtures to simultaneously improve the dimensional tunability and monodispersity in the seeded growth of gold nanorods. *Nano Letters* **2013**, *13*, 765–771.

(33) Jones, M. R.; Mirkin, C. A. Bypassing the limitations of classical chemical purification with DNA-programmable nanoparticle recrystallization. *Angewandte Chemie - International Edition* **2013**, *52*, 2886–2891.

(34) Croft, S. Colloidal gold. *Electron Microscopy Reviews* **2002**, *5*, iii–iv.

(35) Grulke, E. A.; Rice, S. B.; Xiong, J. C.; Yamamoto, K.; Yoon, T. H.; Thomson, K.; Saffaripour, M.; Smallwood, G. J.; Lambert, J. W.; Stromberg, A. J.; Macy, R.; Briot, N. J.; Qian, D. Size and shape distributions of carbon black aggregates by transmission electron microscopy. *Carbon* **2018**, *130*, 822–833.

(36) Grulke, E. A.; Wu, X.; Ji, Y.; Buhr, E.; Yamamoto, K.; Woong Song, N.; Stefaniak, A. B.; Schwegler-Berry, D.; Burchett, W. W.; Lambert, J.; Stromberg, A. J. Differentiating gold nanorod samples using particle size and shape distributions from transmission electron microscope images. *Metrologia* **2018**, *55*, 254–267.

(37) Hu, M. K. Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory* **1962**, *8*, 179–187.

(38) MacSleyne, J. P.; Simmons, J. P.; De Graef, M. On the use of 2-D moment invariants for the automated classification of particle shapes. *Acta Materialia* **2008**, *56*, 427–437.

(39) Park, C.; Huang, J. Z.; Ji, J. X.; Ding, Y. Segmentation, Inference and Classification of Partially Overlapping Nanoparticles. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2012**, *35*, 1–1.

(40) Laramy, C. R.; Fong, L.-k.; Jones, M. R.; Brien, M. N. O.; Schatz, G. C.; Mirkin, C. A. Understanding nanoparticle-mediated nucleation pathways of anisotropic nanoparticles. *Chemical Physics Letters* **2017**, *683*, 389–392.

(41) O'Brien, M. N.; Jones, M. R.; Kohlstedt, K. L.; Schatz, G. C.; Mirkin, C. A. Uniform

circular disks with synthetically tailorable diameters: Two-dimensional nanoparticles for plasmonics. *Nano Letters* **2015**, *15*, 1012–1017.

(42) Sun, S.; Yuan, D.; Xu, Y.; Wang, A.; Deng, Z. Ligand-Mediated Synthesis of Shape-Controlled Cesium Lead Halide Perovskite Nanocrystals via Reprecipitation Process at Room Temperature. *ACS Nano* **2016**, *10*, 3648–3657.

(43) Cho, H.; Shin, J. W.; Ryoo, R.; Ryoo, R. Atomic Scale Mechanisms Underlying Thermal Reshaping of Anisotropic Gold Nanocrystals Revealed by in Situ Electron Microscopy. *Journal of Physical Chemistry C* **2020**, *124*, 12855–12863.

(44) Handwerk, D. R.; Shipman, P. D.; Whitehead, C. B.; Özkar, S.; Finke, R. G. Mechanism-Enabled Population Balance Modeling of Particle Formation en Route to Particle Average Size and Size Distribution Understanding and Control. *Journal of the American Chemical Society* **2019**, *141*, 15827–15839.

(45) Jaggannathan, R.; Mehta, R. V.; Timmons, J. A.; Black, D. L. Anisotropic growth of twinned cubic crystals. *Physical Review B* **1993**, *48*.

(46) Park, K.; Biswas, S.; Kanel, S.; Nepal, D.; Vaia, R. A. Engineering the optical properties of gold nanorods: Independent tuning of surface plasmon energy, extinction coefficient, and scattering cross section. *Journal of Physical Chemistry C* **2014**, *118*, 5918–5926.