SAGE: A Machine Learning Model for Primary Particle Segmentation in TEM Images of Soot Aggregates

Timothy P. Day[a], Khaled Mosharraf Mukut[a], Luke Klacik[a], Ryan O'Donnell[a],
James Wasilewski[a], Somesh P. Roy[a,*]
[a] *Marquette University, Milwaukee, WI, United States*

Abstract

Accurate characterization of the morphology of soot is essential for our understanding and better modeling of the physical and chemical properties of soot. The morphological characteristics of soot are traditionally explored experimentally via Transmission Electron Microscopy (TEM), usually by investigating the images via manual segmentation, which is highly labor intensive. To improve this process, a novel model for automatic segmentation of primary particles in TEM images of soot is presented in this work. The goal of the model is to identify and isolate each primary particle from a TEM image of a soot aggregate. The model, titled Soot Aggregate Geometry Extraction (SAGE) employs a two-stage training process using a convolutional neural network: an initial training on synthetically-generated TEM images followed by a refinement training by using manually segmented real TEM images. The model was tested against a dataset of real TEM images that included images from sources different from the training data (i.e., different instruments and different researchers). When tested against this real TEM image dataset of soot, SAGE shows good performance with an F_1 score of 67.7% indicating its ability to correctly identify primary particles while achieving a balanced trade off between missing true particles and detecting false ones. SAGE is able to detect more primary particles with better shape and size alignments with the ground truth data than traditional methods such as circular Hough transform or Euclidean distance mapping methods leading to a much higher mean Intersection over Union score of 62.2%. Unlike most existing approaches that produce circular segmentations and requires image-by-image tuning, SAGE effectively captures irregular particle boundaries without additional adjustments. The particle size distribution obtained from SAGE matches well with the ground truth. The median errors of predictions obtained from SAGE fall below 5% and 1%, respectively, for radius of gyration and fractal dimension of particles.

Keywords: Soot; TEM; Machine Learning; Primary Particles

---

*Corresponding author.

Information for Editors and Reviewers

1) Novelty and Significance Statement

The research presents a new model for automated segmentation of transmission electron microscopy (TEM) images to extract features of primary particles in a soot aggregate using machine learning. The model is trained using a novel multi-step training process that involves both synthetic images and a few hand-segmented images for robustness and increased performance. The model shows very good performance out of the box, much better than commonly used models, in capturing irregular shapes of primary particles across TEM images from multiple sources. This is a significant development since segmentation of TEM images are often tedious and labor-intensive. The tool presented here shows considerable improvement over existing tools for segmentation of TEM images at primary particle level. This tool will help the community in more efficient analysis of microscopic images to extract morphological information of soot aggregates and primary particles.

2) Author Contributions

- TPD: Writing – original draft, Writing – review & editing , Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.
- KMM: Writing – original draft, Writing – review & editing , Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization.
- LK: Data curation.
- RO: Data curation.
- JW: Data curation.
- SPR: Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

1. Introduction

Soot is a particulate matter – a byproduct of the incomplete combustion of carbon-based fuels with far reaching impact on global climate, public health and welfare [1]. During combustion, physical and chemical processes lead to inception of carbonaceous spherules – clusters of aromatic and aliphatic structures [2, 3] – known as soot primary particles (PP) [4]. As these primary particles grow, they eventually aggregate together to form soot aggregates or simply a soot particle [4, 5]. Along with the internal nanostructure, the morphology of a soot particle – i.e., its size, shape, and spatial arrangement of its primary particles – is an important factor affecting its properties [6]. The soot aggregate's morphology plays a large role in determining its radiative potential and influences how soot contributes to global radiative forcing and climate change [1]. The morphology and size of soot aggregate and primary particles change during their evolution during combustion [6, 7]. Typically, soot aggregates range in size from few to few hundreds of nanometers (nm), with primary particles of sizes up to 40 nm [8].

Transmission electron microscopy (TEM) (and high resolution TEM or HRTEM) is one of the most common diagnostic techniques used to analyze and explore the detailed morphology of soot aggregates [9]. Proper evaluation of TEM images of soot requires classification and segmentation (i.e., identification and detection of boundaries) of primary particles as well as the entire aggregates. This classification and segmentation (referred together as segmentation for brevity in this work) is performed primarily by hand, one particle at a time [10]. Due to the labor-intensive and time-consuming nature of this process, quantitative data on soot morphology is limited. This adds to gaps in our knowledge of morphology variation and its effect on soot properties.

Recent developments in machine learning (ML) and computer vision (CV) have allowed researchers to attempt ML-driven automation in segmentation of microscopic images for feature extraction across various fields [11–15]. In particular, a Convolutional Neural Network (CNN) called Mask R-CNN [16] has proven successful for object detection and semantic segmentation tasks. In soot-related works, techniques such as Otsu thresholding [17], k-means segmentation [18], and CNN approaches [19, 20] have been successfully used in the segmentation of entire (not at primary particle level) soot aggregates. However, manual approach remains the most consistently reliable method to accurately identify, segment, and size *primary particles* within a soot aggregate. An excellent review of primary particle segmentation methods and their limitations can be found in [9]. Work has been done to automate the analysis of primary particles, with some methods such as the pair correlation method (PCM) [17] and Euclidean distance mapping (EDM) [21] providing automated size estimates. Methods such as the circular Hough transform (CHT) [10, 19, 22–25], and a combination of watershed segmentation with EDM (EDM-WS) [9, 11] attempt to provide actual segmentations along with size estimates, but are limited in their capability to handle complex aggregates. These methods struggle with segmentation at the primary particle level due to challenges such as low contrast, noise, morphological variation in primary particles, and extensive overlapping within aggregates. Recently, CNN-based approach was used to attempt primary particle-level segmentation of scanning electron microscopic (SEM) images of particulates using DeepParticleNet (DPN) [20]. But it has not been used on TEM images, which – as mentioned earlier – pose challenges due to low contrast, high degrees of overlap between primary particles, and irregular primary particle boundaries. TEM images are two-dimensional projections of three-dimensional aggregates, so inconsistent contrast makes it difficult to delineate the boundaries between primary particles in densely packed overlapping regions.

Therefore, to achieve segmentation of TEM images at a primary particle level by a ML model, a rigorous training is needed. This introduces a new challenge: To train a ML method, one needs a large amount of quantitative data, which in this case means a large amount of pre-segmented soot TEM images, which is not readily available. In this work, we propose a novel two-phase training approach: A large number of synthetically generated TEM-like images in the first phase followed by a few manually-segmented real TEM images in the second phase. We show that this two-phase training process leads to an improved ML model, titled Soot Aggregate Geometry Extraction (SAGE), capable of detecting primary particles in a soot TEM image. The SAGE model is tested against real TEM images obtained from sources different than the training images to show the model's performance and robustness.

2. Methodology

SAGE is developed by leveraging the instance segmentation capabilities of Mask R-CNN [16], which is a deep convolutional neural network originally developed for object detection and instance segmentation tasks. It builds upon Faster R-CNN [26] by introducing a parallel 'mask head' branch that outputs binary masks for each identified object at the pixel level. This makes it particularly suited for segmentation of densely packed or overlapping primary particles.

SAGE is trained using a two-phase strategy: (i) initial training with synthetically generated TEM images and (ii) fine-tuning with manually annotated real TEM data. Figure 1 illustrates the overall training pipeline. The model is initialized with a standard image detection dataset (COCO [27]) and then first trained on synthetically generated soot TEM images using Mask R-CNN leading to the first variant of the model: SAGE_0. Then it is fine-tuned in two phases using manually annotated real TEM images segmented by different analysts, which leads to two other variants of the model: SAGE_1 and SAGE_2, resulting in progressively refined segmentation performance. Only key stages are shown in Fig 1 for brevity.

2.1. Data preparation and training

To overcome the scarcity of annotated soot TEM data, synthetic aggregates were generated using two software tools: `FLAGE` [28] and `PAGenT` [29]. These tools simulate fractal-like aggregates using particle-cluster algorithms parameterized by primary particle radius, fractal dimension, and overlap ratios. These synthetically generated soot particles were rendered using `PAGenT`'s post-processing abilities to a TEM style image. Gaussian noise was added to mimic imaging artifacts from TEM (see examples in the Supplementary Materials). Each synthetic image was paired with corresponding binary masks of individual primary particles, enabling supervised training. Since these soot particles and their corresponding images are artificially generated to mimic real TEM images, we labeled them as *synthetic TEM images* or *synthetic dataset*.

A limited set of real TEM images [30, 31] was annotated manually using image processing software `ImageJ` [32]. To address subjectivity in manual segmentation, each image was analyzed independently by two randomly assigned analysts from a pool of three, producing two different datasets: M1 and M2. This approach enhances generalization by exposing the model to natural variations in interpretation of primary particle boundary. Table 1 summarizes the datasets used in training.

Table 1: Summary of datasets used for training the model. 'Calc.' segmentation method indicates precalculated PP boundaries for synthetic images. The last column lists the ratio of the dataset used in training and validation.

Dataset	Source	Seg. Method	# of images	Train/ Validate
Synthetic	[28, 29]	Calc.	200	75/25
M1	[30, 31]	Manual	17	80/20
M2	[30, 31]	Manual	17	80/20

Mask R-CNN was initialized with COCO-pretrained weights [27]. It was then trained on the synthetic dataset, which included 200 synthetic images with a total 8,991 primary particles. Loss metrics were monitored during training to mitigate overfitting or underfitting. This resulting ML model obtained with this training is the first variant of the model, labeled as SAGE_0 in this work.

The SAGE_0 model was fine-tuned sequentially using M1 and M2 – the two manually segmented real TEM image datasets. This led to two other variants of the model: SAGE_1 (i.e., SAGE_0 trained with M1) and SAGE_2 (i.e., SAGE_1 trained with M2). This fine-tuning process helped the model learn both general structural patterns and subjective boundary nuances from real TEM data.

2.2. Model performance assessment metrics

The performance of ML models are commonly evaluated using a confusion matrix which is generated by comparing the predicted labels (i.e., primary particle or background) with the ground truth labels. The confusion matrix is composed of four categories, (*a*) *TP* or true positive, i.e., true primary particles, (*b*) *FP* or false positive, i.e., false primary particles, (*c*) *TN* or true negative, i.e., true background, and (*d*) *FN* or false negative, i.e., false background. It is conventional to define metrics such as *accuracy*, *precision*, and *recall* based on these values from confusion matrix [33–35]. As a first order assessment, accuracy tells us how many of the predicted labels are correctly classified as either primary particles or background, whereas precision tells us the proportion of predicted primary particles that are correct, and recall provides an assessment of how many actual (i.e., ground truth) primary particles are predicted. A composite metric, F_1 score, can be formulated to combine both precision and recall, and is given by Eqn. (1).

$$F_1 = \frac{TP}{FP + \frac{1}{2}(FP + FN)} \tag{1}$$

In addition to evaluating classification performance through metrics like accuracy, precision, and recall, we assess segmentation quality using the Intersection over Union (IoU) metric. IoU quantifies the spatial overlap between the predicted segmentation mask and the ground truth mask of a primary particle [35]. It is formally defined as Eqn. (2)

$$\mathrm{IoU} = \frac{|\mathcal{P} \cap \mathcal{G}|}{|\mathcal{P} \cup \mathcal{G}|} \tag{2}$$

where $\mathcal{P}$ is the set of pixels in the predicted mask, and $\mathcal{G}$ is the set of pixels in the ground truth mask.

A higher IoU indicates better alignment between the predicted and true particle shapes. In our evaluations, a threshold of $\mathrm{IoU} > 0.5$ is considered indicative of a "good" segmentation [36]. This metric is particularly important for soot particles, where overlapping and irregular morphologies often challenge instance segmentation models.

None of these metrics alone can provide a comprehensive assessment of the performance of the model and, therefore, they should be considered holistically. These context-agnostic metrics, by themselves, do not capture how well the actual morphology of the aggregate is captured. Therefore, along with the ML-specific metrics, one must also look at domain-specific metrics such as size distribution of primary particles, radius of gyration (R_g) and fractal dimensions (d_f) in evaluating the model performance. The R_g of an aggregate with N primary particles is first calculated in 2D (R_g^{2D}) using the center of each identified particle, (x_i, y_i), and the location of aggregate centroid, $(\bar{x}, \bar{y})$, using Eqn. (3). The R_g^{2D} is then converted to 3D radius of gyration (R_g^{3D}) using Sorenson

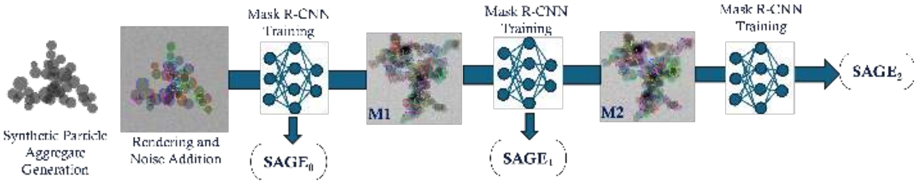


Fig. 1. Training pipeline for SAGE model.

and Feke's scaling factor (Eqn. (4)) for soot with fractal dimensions near 1.8 [37].

$$R_g^{2D} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(|x_i-\bar{x}|^2+|y_i-\bar{y}|^2)} \tag{3}$$

$$R_g^{3D} = 1.023 \times R_g^{2D} \tag{4}$$

A 2D fractal dimension (d_f^{2D}) is calculated via box-counting method [38] using `StereoFractAnalyzer` [39]. The 2D fractal dimension is converted to the commonly used 3D mass fractal dimension (d_f^{3D}), using the correlation shown in Eqn. (5) as proposed in [40, 41].

$$d_f^{3D} = 1.391 + 0.01e^{2.164 d_f^{2D}} \tag{5}$$

3. Results and discussion

The machine learning model developed in this work is evaluated using both ML-specific performance metrics and domain-specific metrics. As mentioned earlier, the ML-specific metrics provide an objective assessment of accuracy of the task performed by the model without considering any contextual information of the image, whereas domain-specific metrics evaluate how well the model predicts various physical properties such as particle size distribution, radius of gyration, and fractal dimension. To provide a baseline, we also compare the performance of our model against two other methods used in the literature, CHT [23, 24] and EDM-WS [11]. These methods were selected for comparison due to their use in the literature to automatically estimate primary particle sizes and approximate their locations and boundaries within aggregates [9]. For both CHT and EDM-WS, we adapted implementations provided in the `atems` (MATLAB Analysis tools for TEM images of Soot) toolbox [42]. Each method was applied to our test image set, and the resulting segmentations were exported as binary masks for comparison with our model, SAGE. For ML-specific metrics and segmentation quality assessment (Sec. 3.1 and 3.2), we performed pixel-by-pixel assessment using the actual masks generated by CHT, EDM-WS, and SAGE. For EDM-WS methods, it is customary to use maximal grey value of EDM raster from EDM-WS method as primary particle sizing parameter. Therefore, for evaluation of the primary particle diameter (d_p), we used the actual diameter of CHT masks, maximal grey value of EDM raster from EDM-WS method, and projected area equivalent diameter of SAGE masks. For radius of gyration and fractal dimension calculations, the circular CHT masks, circular masks based on EDM raster-based d_p, and actual SAGE masks were used. The difference in accuracy of EDM-WS using either the actual watershed masks or using the EDM raster is minimal and the comparison is shown in the Supplementary Materials.

Table 2: Performance of segmentation methods tested on the test dataset. Higher numbers indicate better performance.

	Acc.	F_1	AP_{50}	mean IoU
CHT	14.6%	25.5%	14.0%	29.8%
EDM-WS	28.8%	44.7%	31.7%	40.4%
SAGE_2	51.1%	67.7%	47.4%	62.2%

3.1. ML performance metrics

As the first step, we first evaluated performance of SAGE_0 against a test dataset containing only synthetic TEM images to prove the merits of using a synthetic images in training. In these proof-of-concept tests (see Supplementary Materials), SAGE_0 produces 75.9% accuracy, and an F_1 score of 86.3%. The average IoU in these tests were found to be 92.8%, indicating a very good match with the ground truth. Another metric related to quality of segmentation can be calculated via the average precision at an IoU threshold of 50% (AP_{50}) [27]. The AP_{50} of SAGE_0 against synthetic data was found to be 76.3%.

Next, all the three variants of the model (i.e., SAGE_0, SAGE_1 and SAGE_2) were tested against a test dataset comprised of three new TEM images from the same source as the TEM images used in training [30, 31], and six images from a different source [43–45]. This test dataset includes aggregates of varying sizes and complexities taken in different instruments by different research groups. The use of images from multiple sources in testing helps us determine how well each model can generalize across diverse sets of images. These test images were also analyzed using CHT and EDM-WS to compare to our model. All results and discussion in the rest of this document is related to this real TEM image test dataset. For brevity, only the results from our best model, SAGE_2, is shown here. Results from SAGE_0 and SAGE_1 are shown in Supplementary Materials.

Table 2 compares the segmentation performance of

$SAGE_2$ model with the performances of predictions made with CHT and EDM-WS. Although these metrics are most commonly used in ML applications, they can also be applied to CHT and EDM-WS due to their ability to create instance segmentations. CHT and EDM-WS show relatively low performance across all metrics, with accuracies of 14.6% and 28.8%, respectively. In contrast, $SAGE_2$ performed substantially better, making predictions with 51.1% accuracy. Similar trends are observed in the F_1 scores, which reflect the balance between precision and recall. As with accuracy, $SAGE_2$ achieves the highest F_1 score of 67.7%, indicating its ability to correctly identify particles while achieving a better trade off between missing true particles and detecting false ones. This is much higher than the scores obtained from CHT and EDM-WS, implying that they both miss true particles and detect false ones at higher rates. The AP_{50} scores saw strong gains as well: $SAGE_2$ segmented primary particles with an AP_{50} of 47.4%, far outperforming both CHT and EDM-WS predictions. The mean IoU improved considerably as well, indicating more accurate boundary alignment between prediction and ground truth particles. As a reference, the subjectivity of a human analyst can be measured via the AP_{50} and mean IoU between M1 and M2, which are found to be 68.5% and 66.9%, respectively.

It is noted that CHT's performance can be improved at the cost of generalizability by fine tuning its parameters on an image-by-image basis. On the contrary, once trained, $SAGE_2$ requires no further image-by-image adjustment, even when it is applied to different sized aggregates with different characteristics.

3.2. Segmentation quality

The quality of segmentations made by a model can be further evaluated through IoU distributions – the higher the IoU, the better the segmentation quality (i.e., better match with shape, size, and location of the primary particle). In general, an IoU above 0.5 is considered as an acceptable segmentation [36]. Figure 2 compares the *reverse* cumulative IoU distributions between $SAGE_2$ model and the two other methods, tested against the nine manually segmented test images mentioned previously. In the case of EDM-WS, actual masks created from the watershed segmentation were used to determine IoU. In a reverse cumulative distribution plot, the curve shows how many segmentations have an IoU that is equal to or *above* the IoU threshold on the x-axis.

Comparing the IoU distributions of these methods in Fig. 2, we see that our model makes significantly more segmentations with higher IoUs. For example, $SAGE_2$ detects approximately 112 particles with IoU scores above 0.5, whereas the EDM-WS and CHT only detect 60 and 32 particles, respectively. In fact, there are barely any segmentations made by the other methods with an IoU over 0.7, whereas $SAGE_2$ segments approximately 75 particles with an IoU of 0.7. Therefore, the overall quality of $SAGE_2$'s segmentations match the manual segmentations of the test dataset to a better degree than the two existing methods, highlighting the benefit of training with a mix of synthetic and manually annotated TEM images.

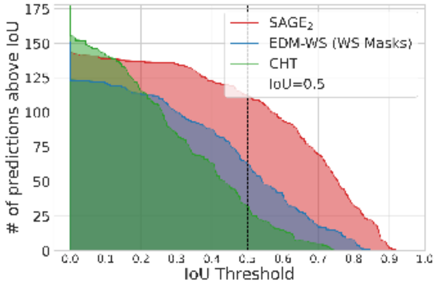


Fig. 2. Reverse cumulative distribution of IoU values for segmentations produced by $SAGE_2$, CHT, and EDM-WS. This distribution shows the count of predictions with IoU values above the threshold on the x-axis.

In addition to the IoU distribution, the quality of segmentations can be visually assessed. Fig. 3 shows the the predicted primary particles from each method on a particular test image. The ground truth manual segmentation is shown in Fig. 3(b). As in the IoU distribution, it is observed that both CHT and EDM-WS struggle to capture the boundaries of primary particles, especially in regions of dense overlap or irregular boundaries. The predictions made by CHT (Fig. 3(b)) look to be significantly smaller than actual primary particle boundaries, and often completely misses detecting particles. EDM-WS (Fig. 3(c)) performs slightly better than CHT and struggles with overlapping particles. $SAGE_2$, albeit not perfect, fares better than the others, making predictions over almost the entire aggregate, capturing irregular boundaries and overlapping primary particles.

3.3. Soot-specific metrics

Morphological information specific to soot aggregates can be extracted from the TEM images by using the segmentations produced by the methods. One of the main morphological characteristics of soot aggregates is the size distribution of its primary particles. The size distributions of the primary particles of the entire test dataset as identified by each segmentation method as well as from the manual method (i.e., the ground truth) are shown in Fig. 4. For EDM-WS, as mentioned earlier, the d_p determined by the EDM, not the projected area equivalent diameter, was used in the size distribution as well as following morphology calculations. As can be seen in Fig. 4, the predicted size distribution from $SAGE_2$ captures the overall spread and shape of the distribution. In contrast, methods like CHT and EDM-WS tend to trend toward smaller diameters as well as less overall detections. The geometric mean primary particle diameter obtainde from $SAGE_2$ is closer to the ground truth than the other two methods. A comparison of the ker-

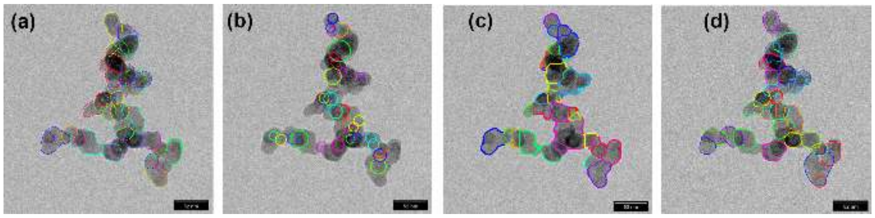


Fig. 3. Primary particles in a sample aggregate from the test set predicted by each segmentation method: (a) manually segmented (ground truth), (b) circular Hough transform, (c) Euclidean distance mapping - watershed, and (d) current SAGE$_2$ Model.

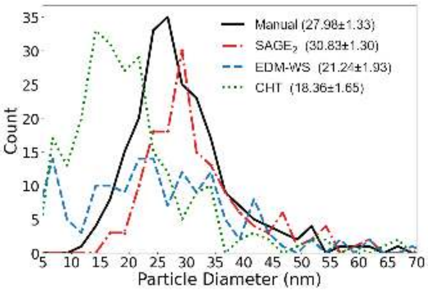


Fig. 4. Primary particle size distribution from the entire test dataset obtained from various methods. The 'Manual' method is the ground truth. The geometric mean and standard deviation of the diameters are shown in the legends.

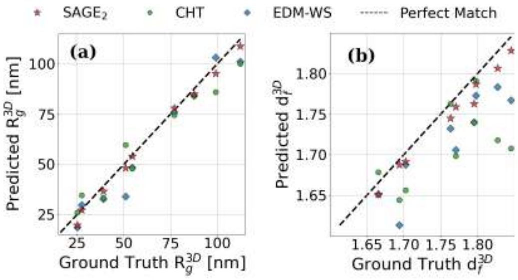


Fig. 5. Comparison of predicted and ground truth morphological metrics. (a) 3D radius of gyration (R_g^{3D}), (b) 3D fractal dimension (d_f^{3D}). The diagonal lines show perfect agreement between prediction and ground truth.

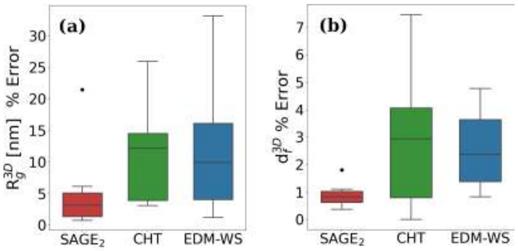


Fig. 6. Absolute % error distribution (ϵ) of (a) predicted 3D radius of gyration (R_g^{3D}) and (b) predicted 3D fractal dimension (d_f^{3D}) with respect to the ground truth dataset.

nel density function showing the same trend is provided in the Supplementary Materials.

Two other soot-specific metrics evaluated in this work – radius of gyration (R_g^{3D}) and fractal dimension (d_f^{3D}) of the aggregates – are shown in Fig. 5, where the predictions from each model is compared against the ground truth. Our SAGE$_2$ model produces morphological measurements that most consistently align with ground truth values, with data points clustering near the line of perfect agreement. In contrast, the segmentations created through CHT and EDM-WS show higher deviations, especially in fractal dimension estimates. This highlights traditional methods' struggles to consistently capture the complexity of various features of aggregates simultaneously. A similar comparison for the 2D variants of these properties (i.e., R_g^{2D} and d_f^{2D}) are provided in the Supplementary Materials showing similar trends.

To further illustrate accuracy, absolute (percent) error (ϵ_Q, where Q can be R_g^{3D} or d_f^{3D}) of the models are calculated from model predictions (Q_m) and the ground truth (Q_t) using Eqn (6).

$$\epsilon_Q = {}^{|Q_m - Q_t|}/_{Q_t} \times 100\% \qquad (6)$$

Fig. 6 presents ϵ_Q distributions for different methods. In both R_g^{3D} and d_f^{3D}, SAGE$_2$ exhibits median errors and spreads far lower than those observed in CHT and EDM-WS. We see that for R_g^{3D}, SAGE$_2$ produces a median error under 5%, with CHT and EDM-WS producing errors exceed 10%. Similarly, the estimated fractal dimensions, d_f^{3D} show a median error below 1% for SAGE$_2$ and around 2-3% for the other methods.

4. Conclusions

This work presents SAGE (Soot Aggregate Geometry Extraction), a machine learning model for segmenting primary particles in TEM soot images. The model is initially trained on synthetic TEM images, then refined using manually segmented real images to better handle complex particle shapes and image textures. Three variants (SAGE$_0$, SAGE$_1$, SAGE$_2$) were developed, with SAGE$_2$ achieving the best performance.

Results show that adding manually-segmented real TEM images in training significantly improves model performance. SAGE$_2$ outperforms CHT and EDM-WS with nearly 3× and 2× higher accuracy respectively, at least 50% higher F_1 scores, and superior AP_{50} and mean IoU. The IoU distribution indicates better shape, size, and location matching with ground truth particles. For soot-specific metrics, $SAGE_2$ accurately reproduces particle size distributions with median errors under 5% for radius of gyration and 1% for fractal dimension. It should be noted that the results are based on a set of nine real TEM images; performance metrics may vary slightly with a larger and more diverse test dataset.

Among existing methods, CHT is limited to circular segmentations and requires image-by-image tuning; while EDM-WS can identify irregular particle shapes but struggles with overlapping primary particles. In contrast, $SAGE_2$ captures irregular, non-circular particle shapes without requiring per-image tuning, making it a robust and efficient solution for automated TEM image segmentation and primary particle feature extraction. Extracting primary particle morphologies from TEM images presents two challenges: the loss of 3D structural information due to 2D projection, and the presence of overlapping particles. While not without limitations, SAGE attempts to overcome these issues by leveraging synthetic training images with known morphologies. Its performance could be further enhanced through inclusion of more diverse and higher quality synthetic and real TEM images of soot.

While some manual effort (in terms of manually segmented real images) was needed in creating the model, once created SAGE does not require any manual intervention or fine tuning and is able to segment primary particles of different sizes obtained from different sources with reasonable accuracy. The trained SAGE models and associated data can be obtained from `www.github.com/comp-comb/SAGE` for use by the research community. SAGE serves as an important step towards greater automation in the segmentation process.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially funded by the Wisconsin Space Grant Consortium (WSGC) through the Dr. Laurel Salton Clark Graduate and Professional Award. The authors acknowledge support from the National Science Foundation as some of this material is based upon work supported by the National Science Foundation under Grant No. 2144290. The authors also acknowledge Dr. Thomas Dreier from the University of Duisburg-Essen and Dr. Kunfeng Gao from ETH Zurich and Beihang University for providing the TEM images used for manual segmentation and model training; and Dr. Tim Sipkens for helpful discussions and assistance with `atems` software.

Supplementary material

Supplementary material related to this article can be found in the "Supplementary Materials" document.

References

[1] T. C. Bond, S. J. Doherty, D. W. Fahey, P. M. Forster, T. Berntsen, B. J. DeAngelo, M. G. Flanner, S. Ghan, B. Kärcher, D. Koch, S. Kinne, Y. Kondo, P. K. Quinn, M. C. Sarofim, M. G. Schultz, M. Schulz, C. Venkataraman, H. Zhang, S. Zhang, N. Bellouin, S. K. Guttikunda, P. K. Hopke, M. Z. Jacobson, J. W. Kaiser, Z. Klimont, U. Lohmann, J. P. Schwarz, D. Shindell, T. Storelvmo, S. G. Warren, C. S. Zender, Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres 118 (11) (2013) 5380–5552.

[2] K. M. Mukut, A. Ganguly, E. Goudeli, G. A. Kelesidis, S. P. Roy, Internal Structure of Incipient Soot from Acetylene Pyrolysis Obtained via Molecular Dynamics Simulations, Journal of Physical Chemistry A 128 (26) (2024) 5175–5187.

[3] H. Wang, Formation of nascent soot and other condensed-phase materials in flames, Proc. Combust. Inst. 33 (1) (2011) 41–67.

[4] H. A. Michelsen, M. B. Colket, P.-E. Bengtsson, A. D'Anna, P. Desgroux, B. S. Haynes, J. H. Miller, G. J. Nathan, H. Pitsch, H. Wang, A review of terminology used to describe soot formation and evolution under combustion and pyrolytic conditions, ACS Nano 14 (10) (2020) 12470–12490, pMID: 32986401.

[5] M. L. Eggersdorfer, S. E. Pratsinis, Agglomerates and aggregates of nanoparticles made in the gas phase, Adv. Powder Technol. 25 (1) (2014) 71–90.

[6] M. J. Thomson, Modeling soot formation in flames and reactors: Recent progress and current challenges, Proc. Combust. Inst. 39 (1) (2023) 805–823.

[7] M. Kholghy, M. Saffaripour, C. Yip, M. J. Thomson, The evolution of soot morphology in a laminar coflow diffusion flame of a surrogate for jet a-1, Combust. Flame 160 (10) (2013) 2119–2130.

[8] R. Dastanpour, S. N. Rogak, Observations of a correlation between primary particle and aggregate size for soot particles, Aerosol Science and Technology 48 (10) (2014) 1043–1049.

[9] T. A. Sipkens, A. Boies, J. C. Corbin, R. K. Chakrabarty, J. Olfert, S. N. Rogak, Overview of methods to characterize the mass, size, and morphology of soot, J. Aerosol Sci. 173 (2023) 106211.

[10] M. Altenhoff, S. Aßmann, C. Teige, F. J. Huber, S. Will, An optimized evaluation strategy for a comprehensive morphological soot nanoparticle aggregate characterization by electron microscopy, Journal of Aerosol Science 139 (1 2020).

[11] P.-J. De Temmerman, E. Verleysen, J. Lammertyn, J. Mast, Semi-automatic size measurement of primary particles in aggregated nanomaterials by transmission electron microscopy, Powder Technol. 261 (2014) 191–200.

[12] H. He, H. Xu, Y. Zhang, K. Gao, H. Li, L. Ma, J. Li, Mask r-cnn based automated identification and extraction of oil well sites, International Journal of Applied Earth Observation and Geoinformation 112 (2022) 102875.

[13] Y. Dai, S. Park, K. Lee, Utilizing mask r-cnn for solid-volume food instance segmentation and calorie estimation, Applied Sciences 12 (2022) 10938.

[14] R. Anantharaman, M. Velazquez, Y. Lee, Utilizing mask r-cnn for detection and segmentation of oral diseases, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 2197–2204.

[15] E. A. Holm, R. Cohn, N. Gao, A. R. Kitahara, T. P. Matson, B. Lei, S. R. Yarasi, Overview: Computer vision and machine learning for microstructural characterization and analysis, Metallurgical and Materials Transactions A 51 (12) (2020) 5985–5999.

[16] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN (2018). `arXiv:1703.06870`.

[17] R. Dastanpour, J. M. Boone, S. N. Rogak, Automated primary particle sizing of nanoparticle aggregates by tem image analysis, Powder Technology 295 (2016) 218–224.

[18] T. A. Sipkens, S. Rogak, Technical note: Using k-means to identify soot aggregates in transmission electron microscopy images, Journal of Aerosol Science 152 (2021) 105699.

[19] T. A. Sipkens, M. Frei, A. Baldelli, P. Kirchen, F. E. Kruis, S. N. Rogak, Characterizing soot in tem images using a convolutional neural network, Powder Technology 387 (2021) 313–324.

[20] M. Frei, F. Kruis, Image-based size analysis of agglomerated and partially sintered particles via convolutional neural networks, Powder Technology 360 (2020) 324–336.

[21] A. Bescond, J. Yon, F. X. Ouf, D. Ferry, D. Delhaye, D. Gaffié, A. Coppalle, C. Rozé, Automated determination of aggregate primary particle size distribution by tem image analysis: Application to soot, Aerosol Science and Technology 48 (2014) 831–841.

[22] A. Cabarcos, C. Paz, R. Pérez-Orozco, J. Vence, An image-processing algorithm for morphological characterisation of soot agglomerates from tem micrographs: Development and functional description, Powder Technology 401 (2022) 117275.

[23] S. Kook, R. Zhang, Q. N. Chan, T. Aizawa, K. Kondo, L. M. Pickett, E. Cenker, G. Bruneaux, O. Andersson, J. Pagels, E. Z. Nordin, Automated detection of primary particles from transmission electron microscope (tem) images of soot aggregates in diesel engine environments, SAE International Journal of Engines 9 (2015) 279–296.

[24] P. Verma, E. Pickering, N. Savic, A. Zare, R. Brown, Z. Ristovski, Comparison of manual and automatic approaches for characterisation of morphology and nanostructure of soot particles, Journal of Aerosol Science 136 (2019) 91–105.

[25] I. Grishin, K. Thomson, F. Migliorini, J. Sloan, Application of the hough transform for the automatic determination of soot aggregate morphology, Applied Optics 51 (2012) 610–620.

[26] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, arXiv (Nov. 2013). `arXiv:1311.2524`, `doi:10.48550/arXiv.1311.2524`.

[27] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: common objects in context, `http://arxiv.org/abs/1405.0312` (2014). `arXiv:1405.0312`.

[28] K. Skorupski, J. Mroczka, T. Wriedt, N. Riefler, A fast and accurate implementation of tunable algorithms used for generation of fractal-like aggregate models, Physica A: Statistical Mechanics and its Applications 404 (2014) 106–117.

[29] I. Gray, T. Day, K. M. Mukut, S. Roy, Pagent (particle aggregate generation tool), `https://github.com/comp-comb/PAGenT` (2024).

[30] T. Dreier, Personal communication [november 2024]: Unpublished transmission electron microscopy (tem) images of soot aggregates, institute for Combustion and Gas Dynamics-Reactive Fluids (IVG-RF) and Center for Nanointegration Duisburg-Essen (CENIDE), University of Duisburg-Essen (2024).

[31] X. Mi, A. Saylam, T. Endres, C. Schulz, T. Dreier, Near-threshold soot formation in premixed flames at elevated pressure, Carbon 181 (2021) 143–154.

[32] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, Nih image to imagej: 25 years of image analysis, Nature Methods 9 (2012) 671–675.

[33] A. Tharwat, Classification assessment methods, Applied Computing and Informatics 17 (2021) 168–192.

[34] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation, BMC Genomics 21 (1 2020).

[35] R. Padilla, S. Netto, E. da Silva, A survey on performance metrics for object-detection algorithms, in: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 2020, pp. 237–242.

[36] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes (VOC) Challenge, Int. J. Comput. Vision 88 (2) (2010) 303–338.

[37] C. M. Sorensen, G. D. Feke, The morphology of macroscopic soot, Aerosol Science and Technology 25 (1996) 328–337.

[38] K. Foroutan-pour, P. Dutilleul, D. L. Smith, Advances in the implementation of the box-counting method of fractal dimension estimation, Applied Mathematics and Computation 105 (1999) 195–210.

[39] K. Mukut, Stereofractanalyzer, `github.com/comp-comb/StereoFractAnalyzer` (2024).

[40] C. Lee, T. A. Kramer, Prediction of three-dimensional fractal dimensions using the two-dimensional properties of fractal aggregates (12 2004).

[41] T. Honablew, M. N. Fiddler, R. P. Pokhrel, S. Bililign, Fractal dimensions of biomass burning aerosols from tem images using the box-grid and nested squares methods, Atmosphere 14 (2 2023).

[42] T. A. Sipkens, atems (MATLAB analysis tools for TEM images of soot), `https://github.com/tsipkens/atems` (2020).

[43] K. Gao, F. Friebel, C.-W. Zhou, Z. A. Kanji, Enhanced soot particle ice nucleation ability induced by aggregate compaction and densification, Atmospheric Chemistry and Physics 22 (7) (2022) 4985–5016.

[44] K. Gao, C.-W. Zhou, E. J. B. Meier, Z. A. Kanji, Laboratory studies of ice nucleation onto bare and internally mixed soot–sulfuric acid particles, Atmospheric Chemistry and Physics 22 (8) (2022) 5331–5364.

[45] K. Gao, F. Friebel, C.-W. Zhou, Z. A. Kanji, Quantifying densification effects on the morphology of nano-sized carbon black aggregates, Powder Technology 459 (2025) 120943.

Supplementary Materials for “SAGE: A Machine Learning Model for Primary Particle Segmentation in TEM Images of Soot Aggregates”

Timothy P. Day, Khaled Mosharraf Mukut, Somesh P. Roy[*]

Marquette University, Milwaukee, WI, USA.

Abstract

This document contain additional data supplementary to the original article. The following data and plots are presented here. For acronyms and appropriate references, please see the main text of the article. The SAGE models and associated data can be obtained from `www.github.com/comp-comb/SAGE`.

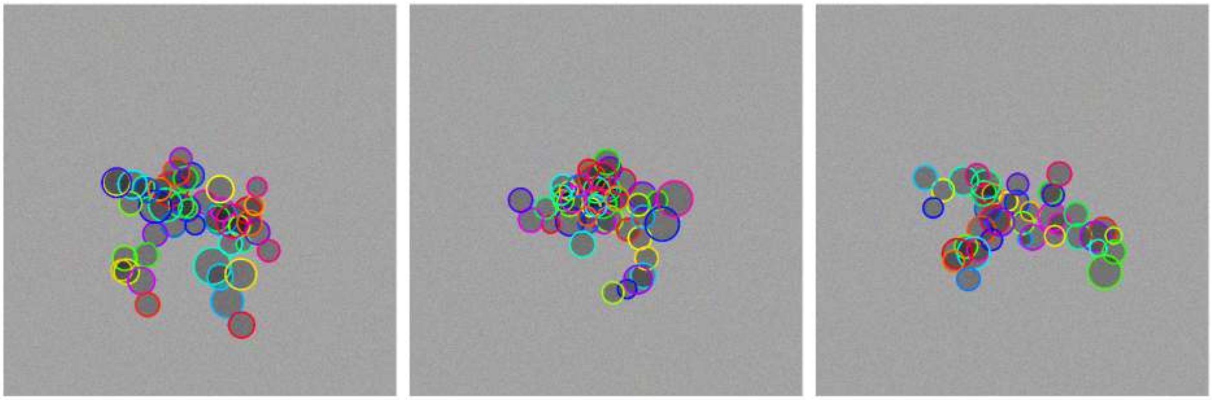
Contents

---

[*] Corresponding author: somesh.roy@marquette.edu

1 Sample synthetic TEM images

A sample of three synthetic TEM images of soot (which are used to train SAGE_0) are shown below in Fig. S1.



Supplementary Figure S1. Examples of synthetically generated TEM images, used in training SAGE_0.

2 Proof-of-concept tests on synthetic image datasets

This section contains performance comparison tables between different variants of SAGE models (i.e., SAGE_0, SAGE_1, and SAGE_2) and models only trained on COCO dataset followed by fine tuning via M1 and M2 dataset (referred as COCO_1 and COCO_2). This data show the improvement of adding synthetic images in training over standard COCO-only training.

Table 1: Performance metrics of SAGE_0 when tested against dataset containing synthetic TEM images only.

Model	**Test set identifier**	**Acc.**	**F_1**	**AP_{50}**	**mean IoU**
SAGE_0	Synthetic-only	75.9%	86.3%	76.3%	92.8%

Table 2: Comparison of performance metrics of different variants of SAGE and COCO-initialized model (i.e., no synthetic image in the training process) when tested against dataset containing real TEM images only. SAGE models perform consistently better than COCO-only models indicating the improvement gained by adding synthetic TEM images in the training process.

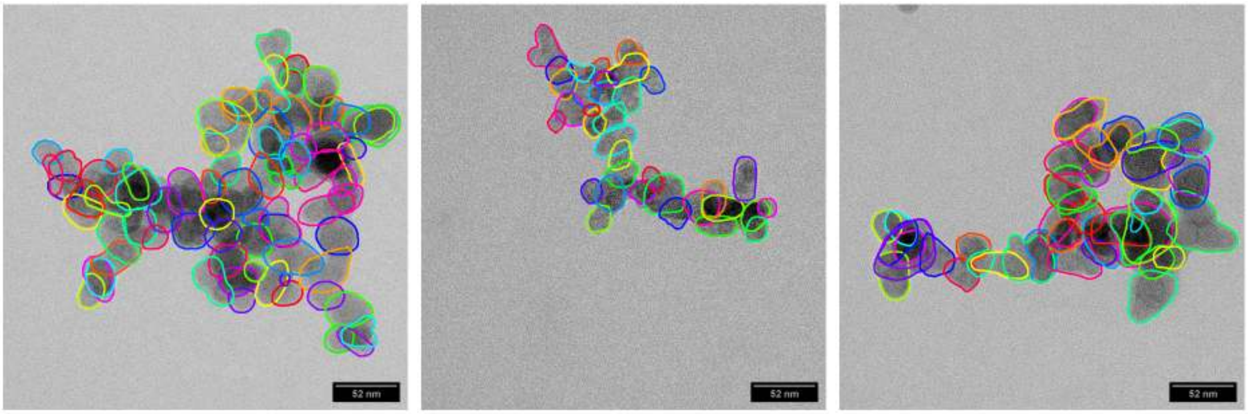
Model	**Test set identifier**	**Acc.**	**F_1**	**AP_{50}**	**mean IoU**
SAGE_0	Real TEM images	57.0%	72.7%	33.4%	50.2%
SAGE_1	Real TEM images	48.0%	64.7%	43.8%	64.4%
COCO_1	Real TEM images	45.2%	62.3%	46.1%	59.0%
SAGE_2	Real TEM images	51.1%	67.7%	47.4%	62.2%
COCO_2	Real TEM images	43.8%	61.0%	39.1%	59.2%

3 Manual segmentation procedure and uncertainty in manual segmentation

This section outlines the general procedure for creating manual segmentation for use in SAGE. Manual segmentation was performed using `ImageJ` [1].

1. In the software `ImageJ`, open a TEM image from preferred file format (.dm3, .tif, etc).
2. Open the 'ROI manager' to save multiple selections.
3. Using the freehand selection tool, outline a primary particle .
4. Add selection to ROI manager, and repeat until all primary particles are outlined. Each ROI is essentially a primary particle.
5. In regions with high levels of overlap, use the "Enhance Local Contrast" plugin to increase contrast to make some overlapping boundaries easier to see. However, avoid increasing contrast too much, or artifacts may appear leading to hallucinated boundaries.
6. With all primary particles in an aggregate saved to ROI manager, use a macro to convert each individual ROI into a binary image mask and save to proper directory.
7. These binary masks are then used as ground truth data for SAGE training, validation, and testing.

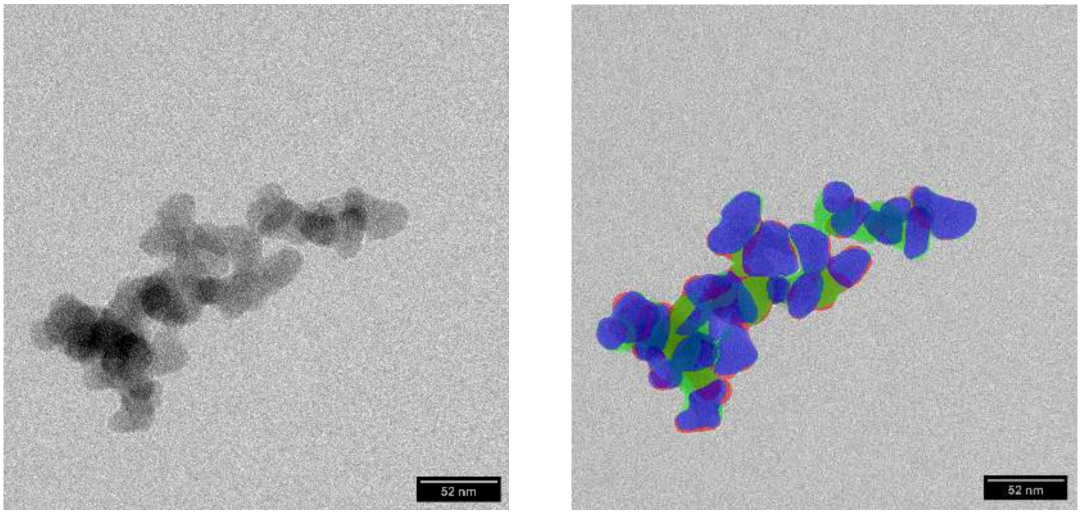
Figure S2 shows examples of manually segmented real TEM images that were used in training SAGE.



Supplementary Figure S2. Examples of manually segmented TEM images of soot, performed using ImageJ

An example of differences rising from segmentation of the same image by two different analysts is shown in Fig. S3. The user error in manual segmentation can be quantified using average precision (AP_{50}) and IoU, which indicates how well one analyst's segmentation match another's. Between the two sets of manually segmented training images, we see an AP_{50} of 68.5% and a mean IoU of 66.9%. The difficulty in manual segmentation is that it can be very subjective based on how different people are discerning primary particle boundaries. It is by no means an exact science, so at no point would it be expected for different analysts to segment aggregates identically. Because SAGE is trained on both sets of segmentations, it is able to learn different viewpoints and potential ways to interpret the images it analyzes.

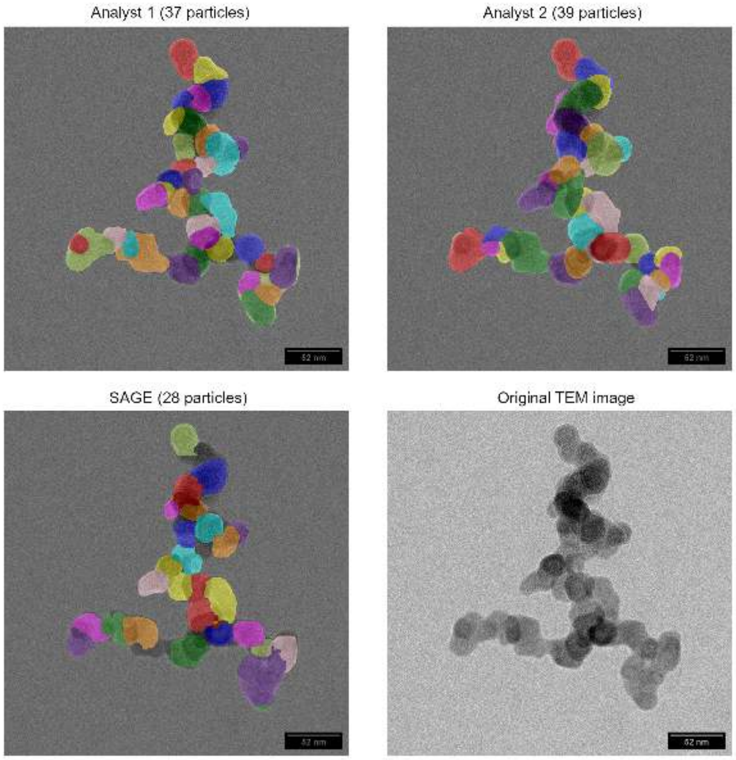
As an example, the differences in the segmentations done manually and by SAGE are shown in Figure S4.



(a) TEM image

(b) Differences in segmentations.

Supplementary Figure S3. Difference in segmentation done by two analysts. Region in red indicates segmentations done only by Analyst 1, green regions show segmentations done only by Analyst 2, and the blue (with 50% opacity) regions indicate the common overlap region between two segmentations.

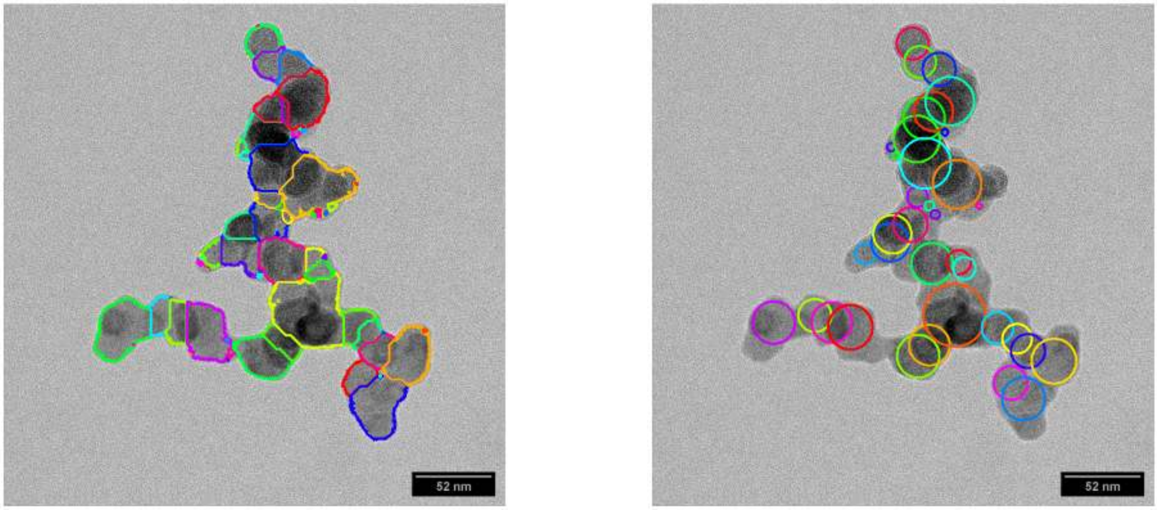


Supplementary Figure S4. An example of segmentations produced by different human analysts and SAGE_2

4 Additional results

4.1 Comparison of two approaches for EDM-WS size estimation

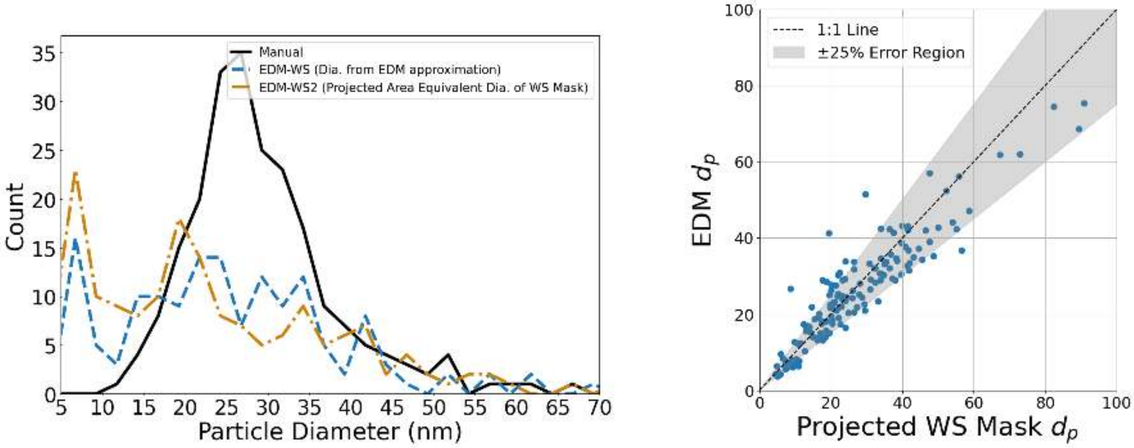
The two approaches for size measurement in EDM-WS method are (1) use the actual area of WS masks and (2) use the maximal grey value of the EDM raster. This subsection compares the results of sizing and morphological evaluations using both approaches.



(a) EDM-Watershed Segmentation Mask

(b) Circular masks created using the maximal grey value of EDM raster

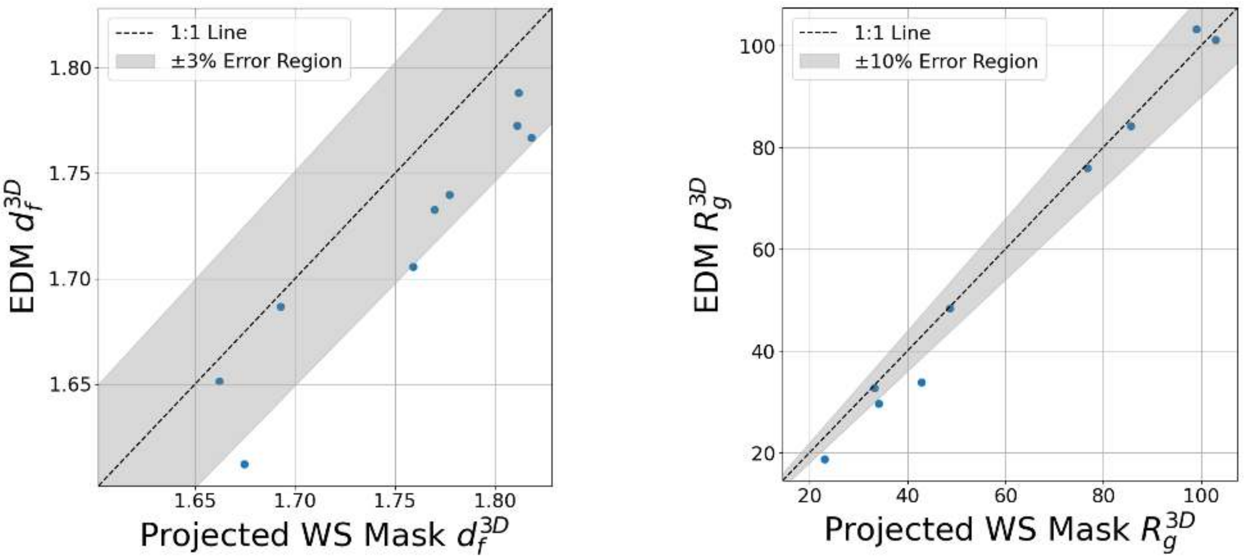
Supplementary Figure S5. Comparison of two size measurement approaches in EDM-Watershed segmentation method.



(a) Primary particle size distribution obtained from both EDM-WS approaches

(b) Comparison of primary particle size obtained from both EDM-WS approaches

Supplementary Figure S6. Comparison of primary particle size obtained from both EDM-WS approaches

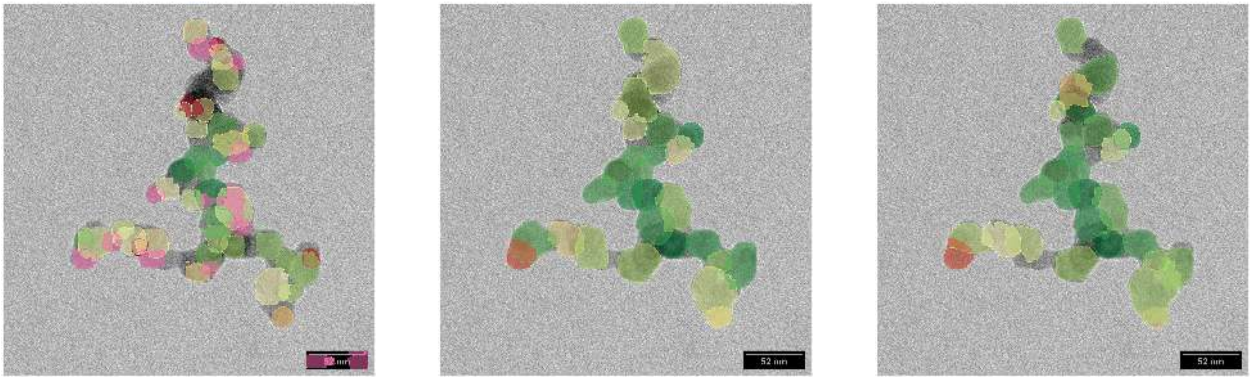


(a) Comparison of 3D fractal dimension obtained from both EDM-WS approaches

(b) Comparison of 3D radius of gyration obtained from both EDM-WS approaches

Supplementary Figure S7. Comparison of morphology metrics obtained from both EDM-WS approaches

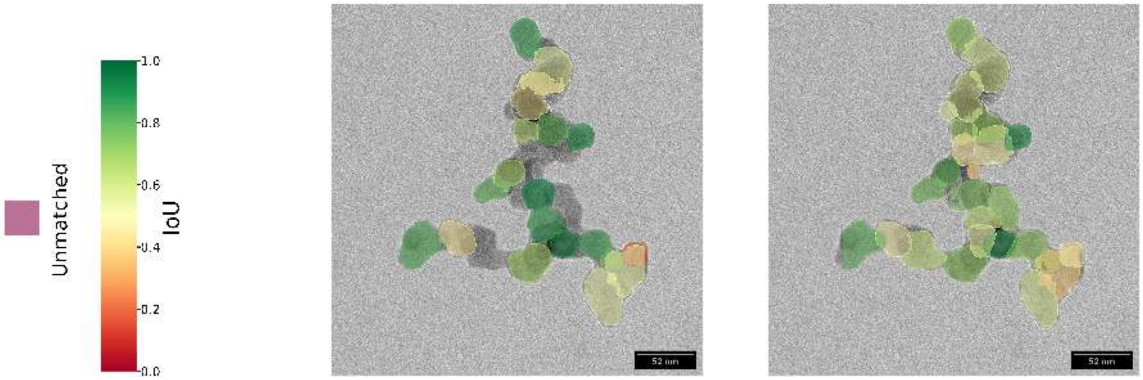
4.2 Sample comparison of IoU heat maps from SAGE and COCO



(a) SAGE_0

(b) SAGE_1

(c) SAGE_2

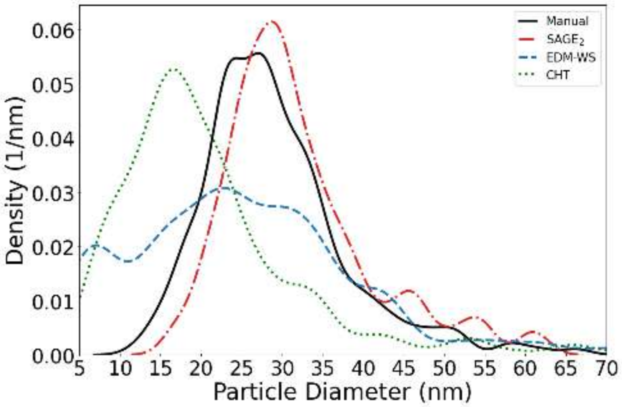


(d) COCO_1

(e) COCO_2

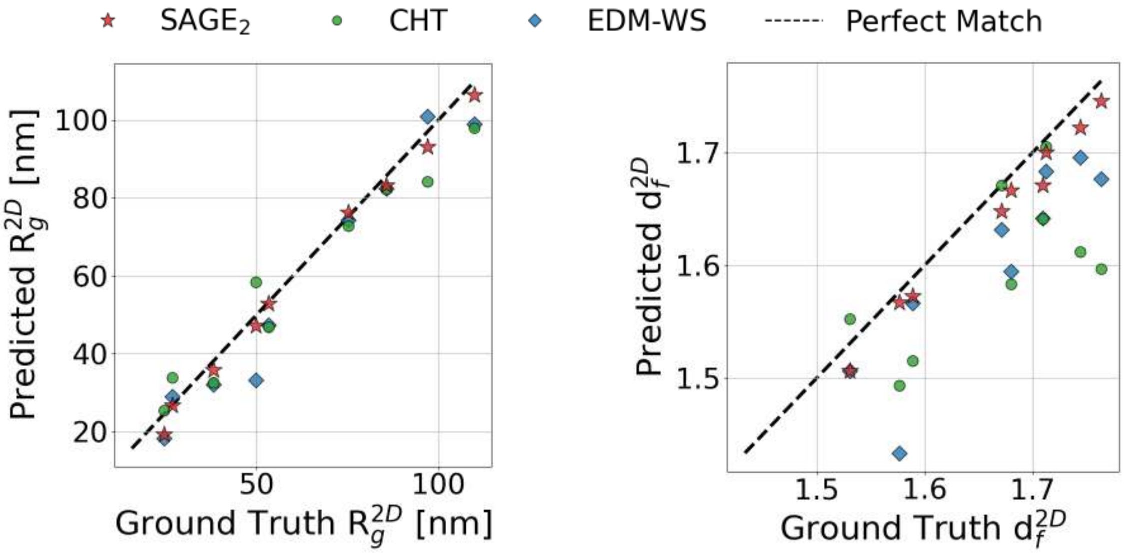
Supplementary Figure S8. IoU mask heat maps for a sample image when tested with different models. The higher the IoU value, the better the segmentation quality. (a) SAGE_0: TP: 24 , FP: 27 mean IoU: 55.9%, (b) SAGE_1: TP:, 24 FP 4, mean IoU: 69.8%, (c) SAGE_2: TP: 24, FP 4, mean IoU: 70.1%, (d) COCO_1: TP: 16 , FP: 3, mean IoU: 68.4%. (e) COCO_2: TP: 20, FP: 6 , mean IoU: 62.4%

4.3 Kernel density estimate of primary particle size



Supplementary Figure S9. Kernel density estimate of particle sizes from the entire test dataset (real TEM images) obtained from various models. The 'Manual' method is considered the ground truth. The maximal grey value of EDM-WS is used in primary particle sizing from EDM-WS method.

4.4 Two-dimensional radius of gyration and fractal dimension



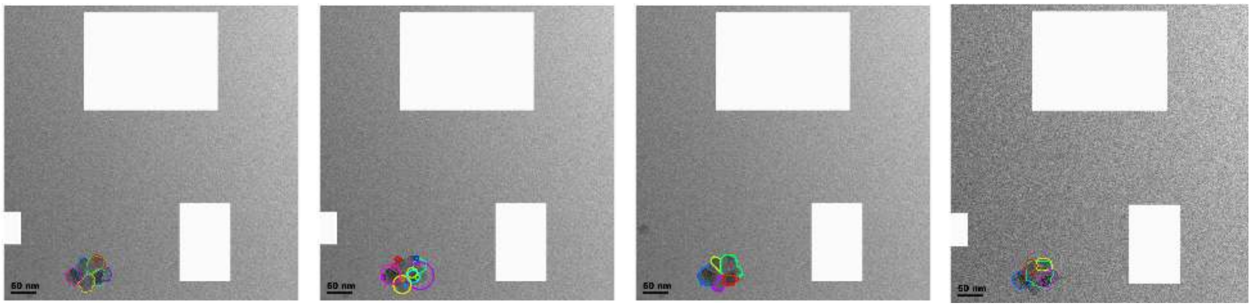
(a)

(b)

Supplementary Figure S10. Comparison of predicted and ground truth 2D morphological metrics across test dataset (real TEM images). (a) Predicted vs. Ground Truth R_g^{2D}. (b) Predicted vs Ground Truth d_f^{2D}. The dashed diagonal lines represent perfect agreement between prediction and ground truth values.

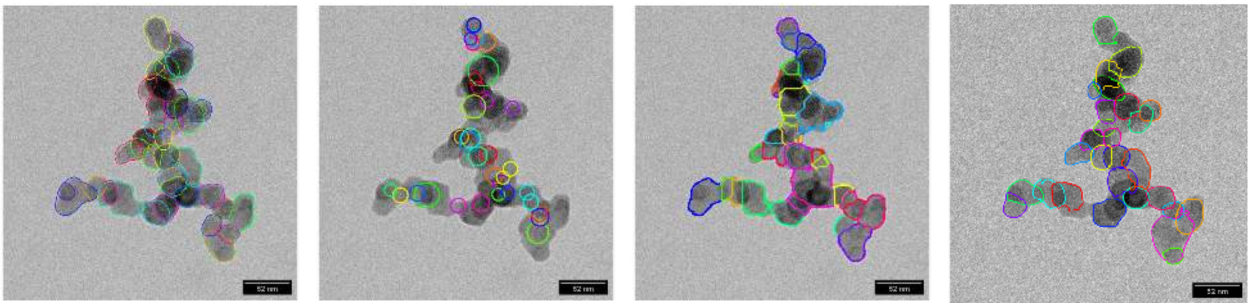
5 Segmentations from all models for entire test dataset

The segmentation results from all nine test images (please see citations in the captions for the source of the original TEM image) including the manual segmentations are presented here. Some TEM images contained more than one aggregate or some other artifacts. These regions were was cropped out (indicated by white boxes) so that only one aggregate is present in each image. Future iterations of SAGE may include support for handling multiple aggregates.



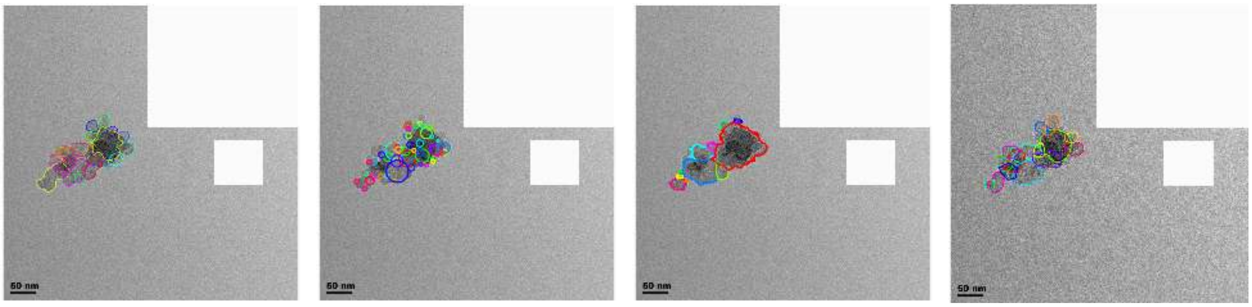
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE_2

Supplementary Figure S11. Segmentation of Test Image #1 [2–4]



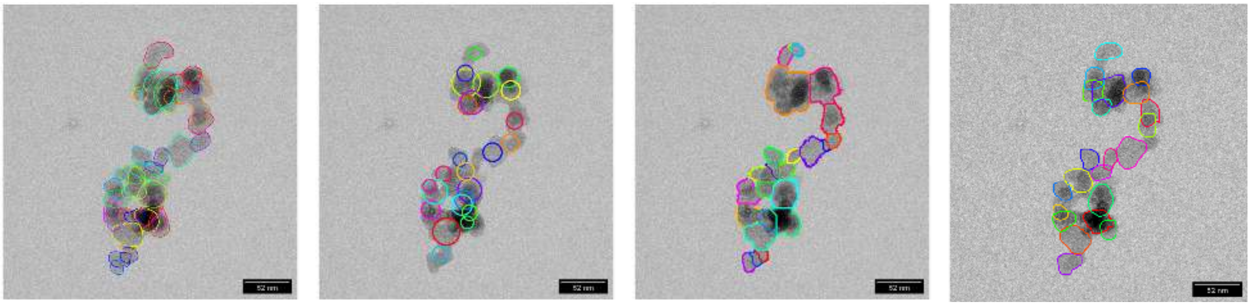
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE_2

Supplementary Figure S12. Segmentation of Test Image #2 [5, 6]



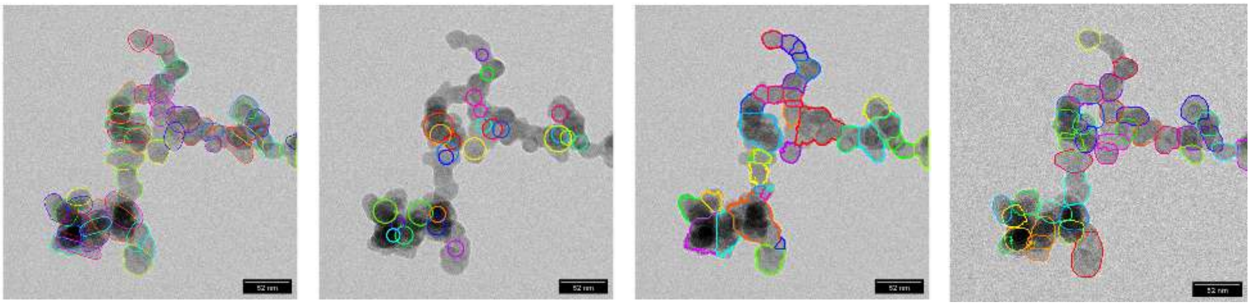
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE_2

Supplementary Figure S13. Segmentation of Test Image #3 [2–4]



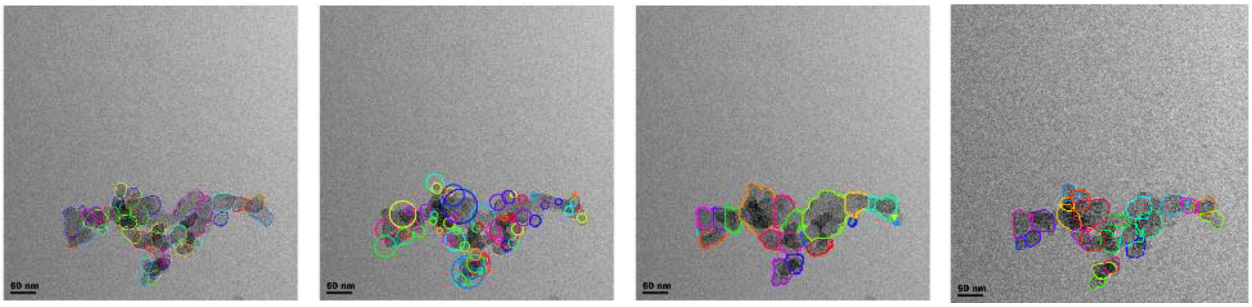
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S14. Segmentation of Test Image #4 [5, 6]



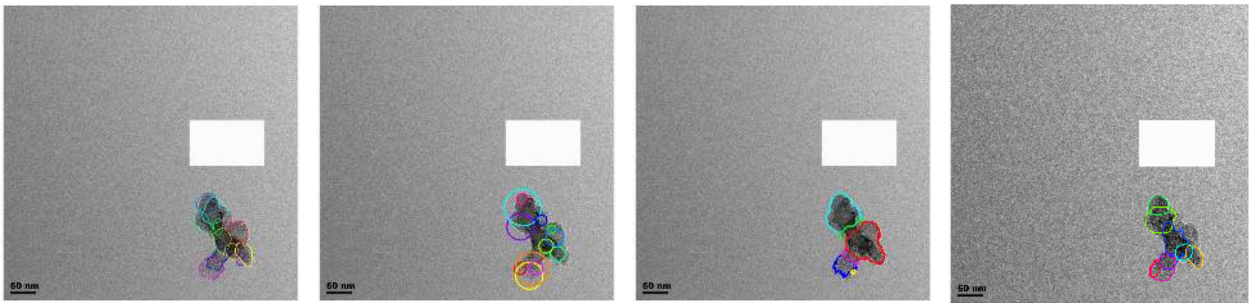
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S15. Segmentation of Test Image #5 [5, 6]



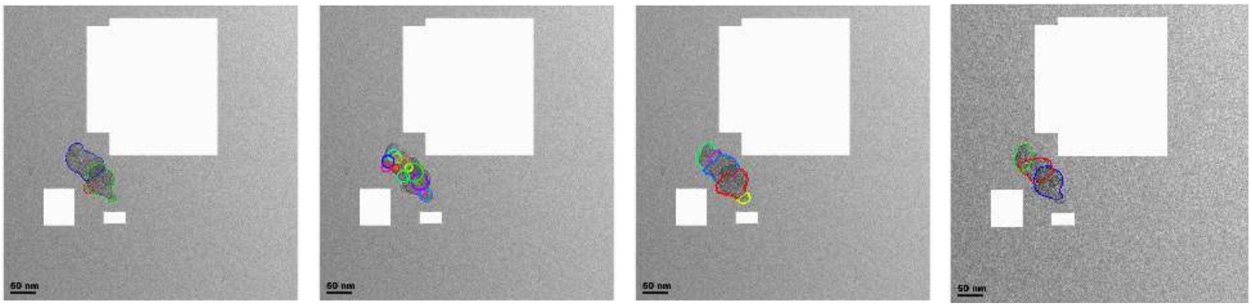
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S16. Segmentation of Test Image #6 [2–4]



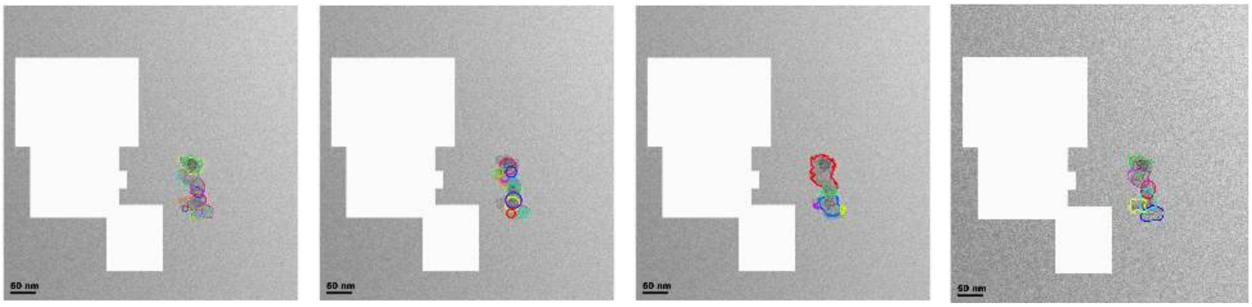
(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S17. Segmentation of Test Image #7 [2–4]



(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S18. Segmentation of Test Image #8 [2–4]



(a) Manual Segmentation (b) CHT (c) EDM-WS (d) SAGE$_2$

Supplementary Figure S19. Segmentation of Test Image #9 [2–4]

References

[1] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, Nih image to imagej: 25 years of image analysis, Nature Methods 9 (2012) 671–675.

[2] K. Gao, F. Friebel, C.-W. Zhou, Z. A. Kanji, Enhanced soot particle ice nucleation ability induced by aggregate compaction and densification, Atmospheric Chemistry and Physics 22 (7) (2022) 4985–5016.

[3] K. Gao, C.-W. Zhou, E. J. B. Meier, Z. A. Kanji, Laboratory studies of ice nucleation onto bare and internally mixed soot–sulfuric acid particles, Atmospheric Chemistry and Physics 22 (8) (2022) 5331–5364.

[4] K. Gao, F. Friebel, C.-W. Zhou, Z. A. Kanji, Quantifying densification effects on the morphology of nano-sized carbon black aggregates, Powder Technology 459 (2025) 120943.

[5] T. Dreier, Personal communication [november 2024]: Unpublished transmission electron microscopy (tem) images of soot aggregates, institute for Combustion and Gas Dynamics-Reactive Fluids (IVG-RF) and Center for Nanointegration Duisburg-Essen (CENIDE), University of Duisburg-Essen (2024).

[6] X. Mi, A. Saylam, T. Endres, C. Schulz, T. Dreier, Near-threshold soot formation in premixed flames at elevated pressure, Carbon 181 (2021) 143–154.