Capstone 3: A successful deep learning story.

Daniel Groom[1]

[1]Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, 78712, USA

I first began studying inorganic nanoparticles during my MS, which focused on their growth mechanisms in the cathodes of proton exchange membrane fuel cells (PEMFCs). My goal was to employ (scanning) transmission electron microscopy ((S)TEM) to study particle size at different conditions to gain a better understanding of what causes nanoparticles to grow. Imaging these nanoparticles in the (S)TEM proved rather straightforward, and it became clear that I could accurately image large numbers of them for measurement through digital image processing. However, I soon learned that no image processing software existed for the task of segmenting and sizing inorganic nanoparticles in (S)TEM images. The sizing had to be done manually or through crude and inaccurate automation assisted routines.

That task quickly became a bottleneck for my studies, so I finished up my MS and moved on to the subject of my PhD, developing software for segmenting nanoparticles in (S)TEM images. At the beginning of this task, I conducted a thorough literature survey and found myself somewhat overwhelmed. The images under consideration in the literature were incredibly simple and unrealistic for what (S)TEM produces for these samples (see Figure 1 for examples). Moreover, the algorithms they discussed were ancient in image processing terms, and they performed poorly on even simple images. So, in response, I invented a novel technique that I called hybrid thresholding. Figure 2 compares two traditional methods to hybrid thresholding, showing that my algorithm performs the best for semantic segmentation.

I also found numerous approaches for instance segmentation in the literature. However, I found that the most reliable method was to segment only isolated particles, excluding the rest of the image (see Figure 3 for a schematic of this paradigm). The most important result of my early studies was to demonstrate that my algorithm reproduced particle size distributions (PSDs), whereas traditional methods did not.

Shortly after I began my studies, the research community introduced deep learning for this task. Deep Learning models have provided state of the art results for other segmentation problems, so I was encouraged at this introduction. Unfortunately, the improvements provided by deep learning models were only marginal, and no researchers had published their models to share with the research community. If I wanted to study deep learning Architectures extensively, I would have to do the hard work myself. That brings us to this study. Having introduced the problems in this section here, I move on to proposed methods and studies in the next section.

II. Proposed methods and studies

We begin this section with an overview of a comprehensive pool of images that sets our study apart from others. This image pool is 1700 images composed of at least 20 images per parameter in Table 1 below.

Validation Parameter	Variables
Particle Size (mean)	1. ~2nm 2. ~3nm 3. ~5nm 4. ~12nm
Particle Composition	1. Pt 2. Pt3Co 3. FeP 4. SnCu
Substrate	1. Graphene 2. Carbon nanotubes 3. Carbon powders 4.PEMFC MEA cathodes
Modalities	1. BFTEM 2. BFSTEM 3. HAADF-STEM 4. LAADF-STEM
Instrument	1. JEOL 2010F in Austin TX 2. JEOL 2100 in Portugal 3. FEI Titan ChemiSTEM in Portugal 4. JEOL 200 in Japan

Table 1. Parameters for our collection of 1700 TEM images

We set aside 150 of these images at random for a comparison database, leaving the remaining 150 images as a deep learning database. For the sake of deep learning, we dedicate 20% of those images to validation and the remaining 80% for training.

The first study I propose is to compare traditional methods (global, local, and hybrid thresholding) to three fully convolutional networks (FCNs), namely the UNET, FPN, and Linknet (all pretrained) How do their performances compare on our comparison database of 150 images?

The next study will be to focus on outliers in the FCNs' performance, aiming to elucidate why the algorithms fail when they do. Specific variables to investigate will be contrast and size. Are these approaches contrast independent? Are they size invariant? The goal of this analysis is to improve the performance of these FCNs by identifying root causes of their performance failures.

III. Results and Discussion

Figure 5 contains the results for the Jaccard similarity coefficient (JSC) distributions for each method under consideration. Without dealing with any summary statistics or formal scoring, it is clear that hybrid thresholding greatly outperforms the other methods. Surprisingly, the deep learning approaches offer approximately the same performance as global thresholding, a highly undesirable result.

To better understand these failures, we first study the impact of contrast on performance. Now, intuitively, the performance of a given method should increase with increasing contrast. And, in the best-case scenario, performance would be essentially contrast independent. Let's see how our algorithms perform with respect to contrast by reviewing Figure 6. Traditional methods perform as expected, and hybrid thresholding even appears to operate independently of contrast. However, the deep learning methods all show inverse directionality with contrast, a counterintuitive and highly undesirable result. To counteract this behavior, we propose adding 100 high contrast images and ground truths to the database.

A second issue to consider is the size dependence of these algorithms as shown in Figure 7. Notice how again the traditional methods perform as one might expect, with hybrid thresholding operating almost independently of size. However, the deep learning methods show strong size dependency, with each method showing a sharp drop off in performance around 45 pixels in area equivalent diameter. To counteract this behavior, we propose changing the kernel size of the deep learning models.

Figure 8 contains the results of hybrid thresholding compared to the UNET with a kernel size of 7 (UNET7) after adding the high contrast images and ground truths to the database. It is clear from this figure that UNET7 greatly outperforms the previous deep learning models. However, it is also clear that UNET7 fails to outperform hybrid thresholding.

IV. Conclusion and future work

In this work, we have demonstrated how to optimize the performance of deep learning models. While the deep learning model we proposed failed to outperform hybrid thresholding, this will not always be the case. Once we add a larger number of images to the training database, the end result should be a powerful deep learning method that outperforms even hybrid thresholding.

To give an idea of the type of performance these approaches can provide, consider Figures 9 and 10. Figure 9 shows the instance segmentation of a HAADF STEM image when hybrid thresholding is employed. Figure 10 shows the instance segmentation of a BF STEM image when hybrid thresholding is employed. In both cases, the segmentation can be seen to be almost perfect. This is the future of segmentation for inorganic nanoparticles in (S)TEM images.

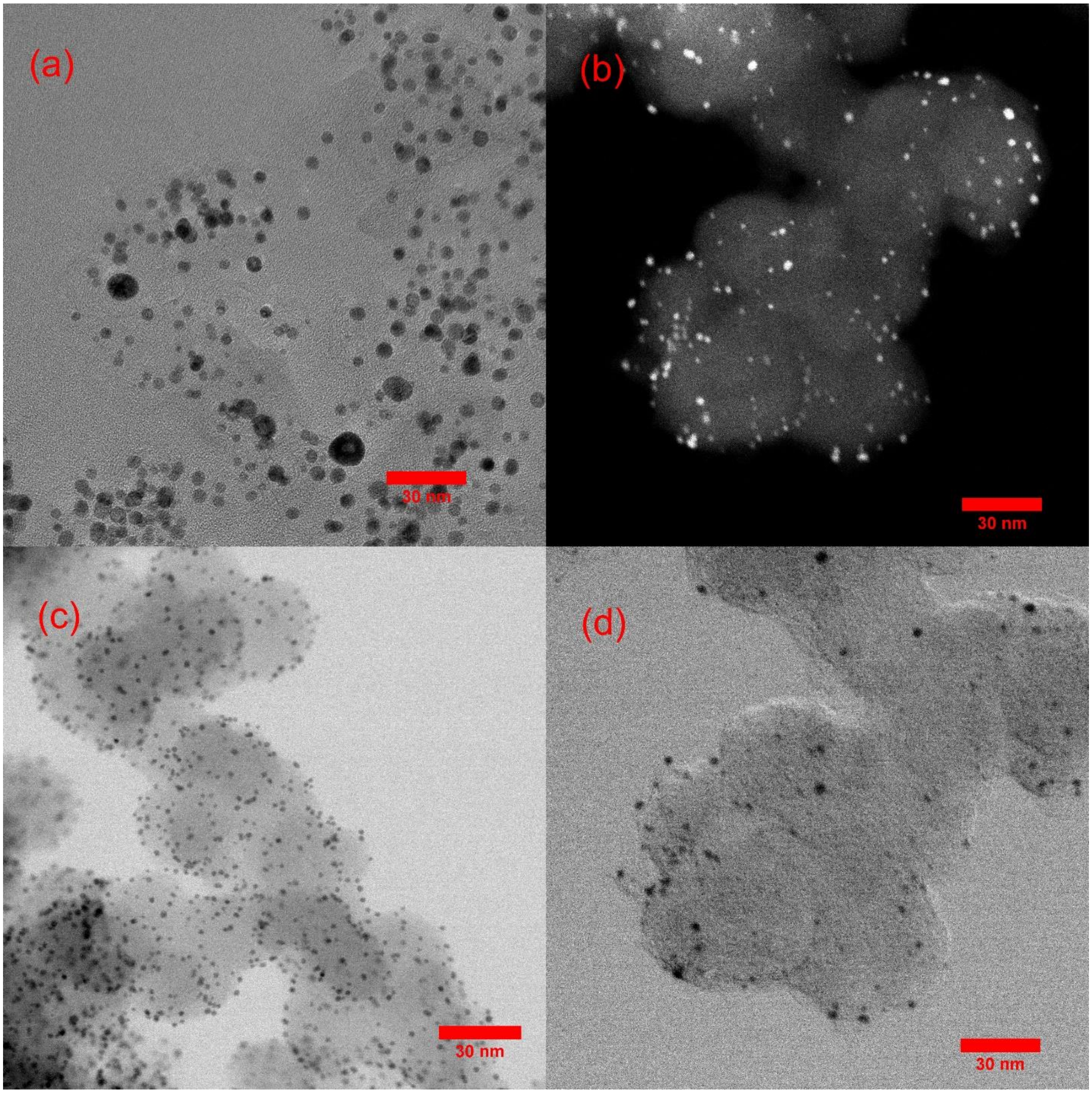

Figure 1.TEM/STEM images using four different modes: (a) Bright-Field TEM (BF TEM); (b) High Angular Annular Dark Field STEM (HAADF STEM); (c) Bright-Field STEM (BF STEM); and (d) Low Angular Annular Dark Field STEM (LAADF STEM).

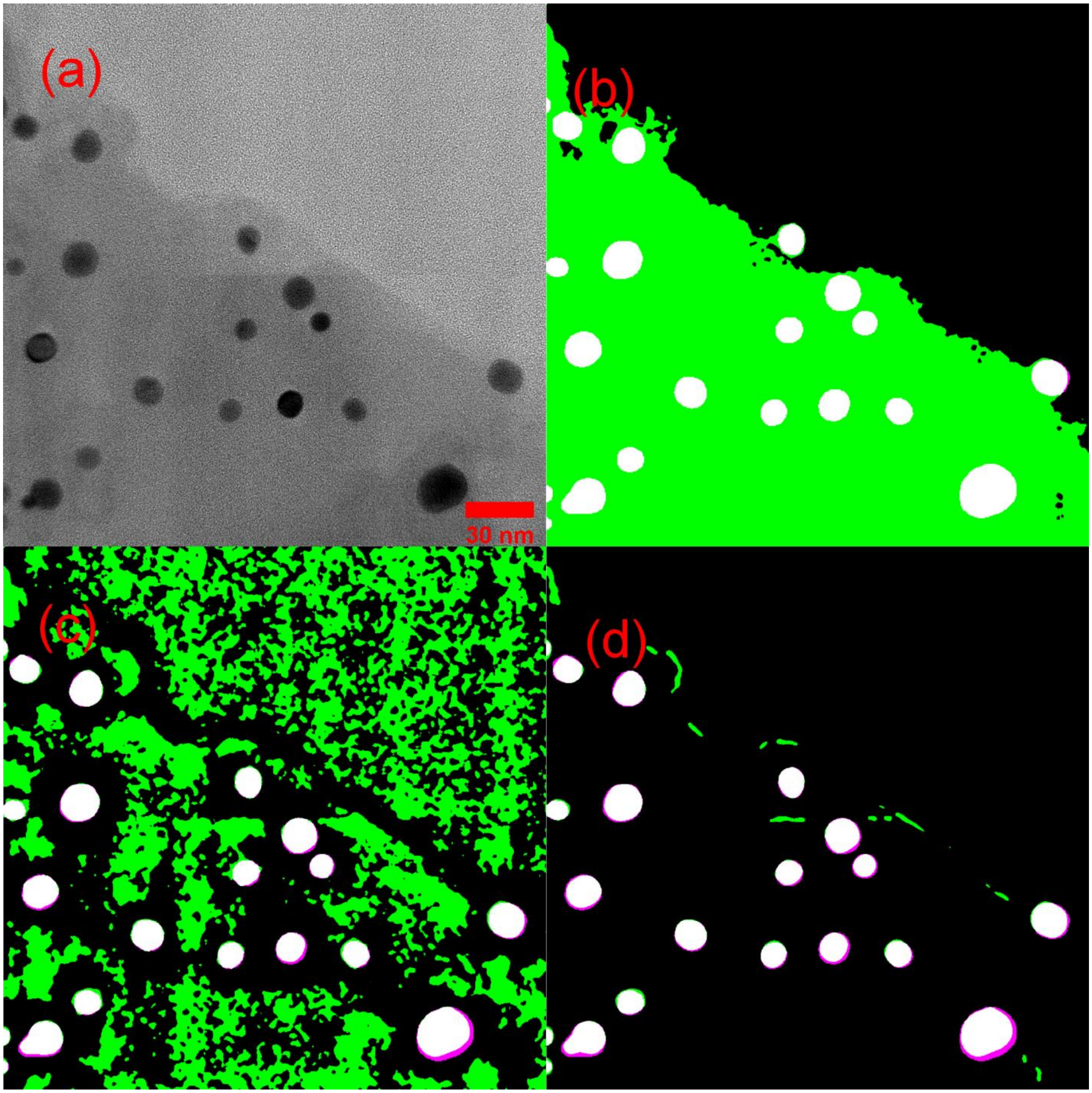

Figure 2. Sematic segmentation results in comparison to ground truth: (a) original BFTEM image; (b) global thresholding; (c) local thresholding; (d) hybrid thresholding. In the images (b) to (d), black and white regions are areas of agreement, whereas colored regions are areas of disagreement.

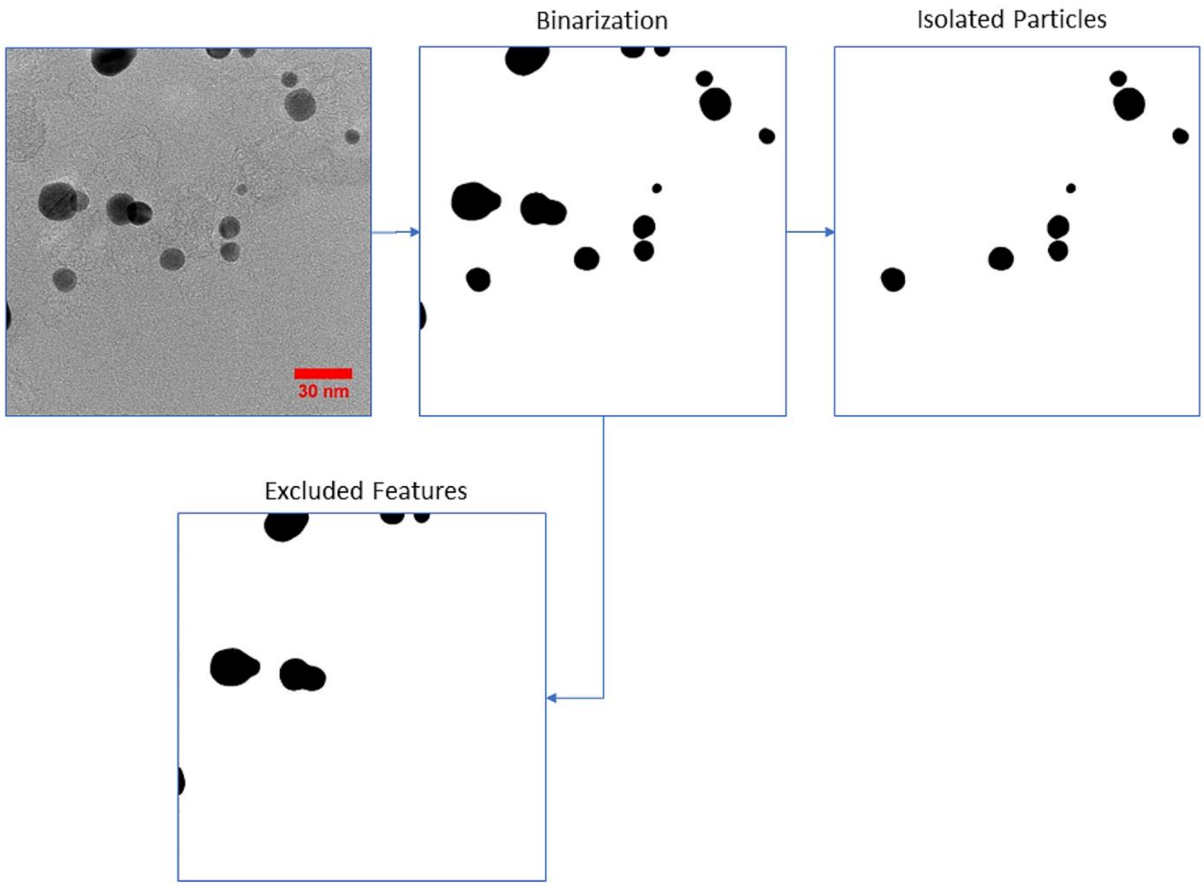

Figure 3. Traditional particle picking paradigm for segmenting isolated and overlapping particles.

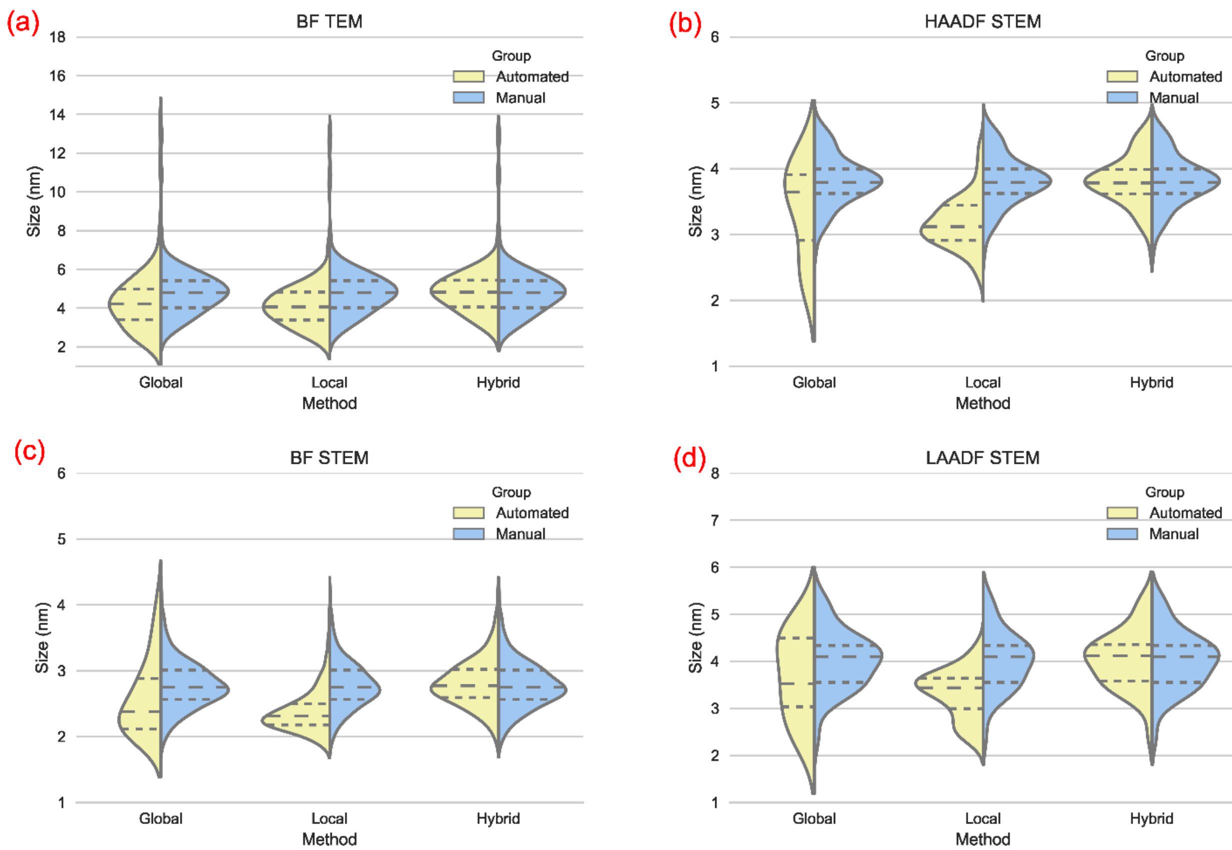

Figure 4. Manual and automated particle sizing results from the images in Figure 1. The four different modalities are (a) BF TEM; (b) HAADF STEM; (c) BF STEM; and (d) LAADF STEM. Automated results for global thresholding, local thresholding, and hybrid thresholding are shown in yellow, whereas manual segmentation is shown in blue. Notice how only hybrid thresholding reproduces manual particle sizing for all four modalities.

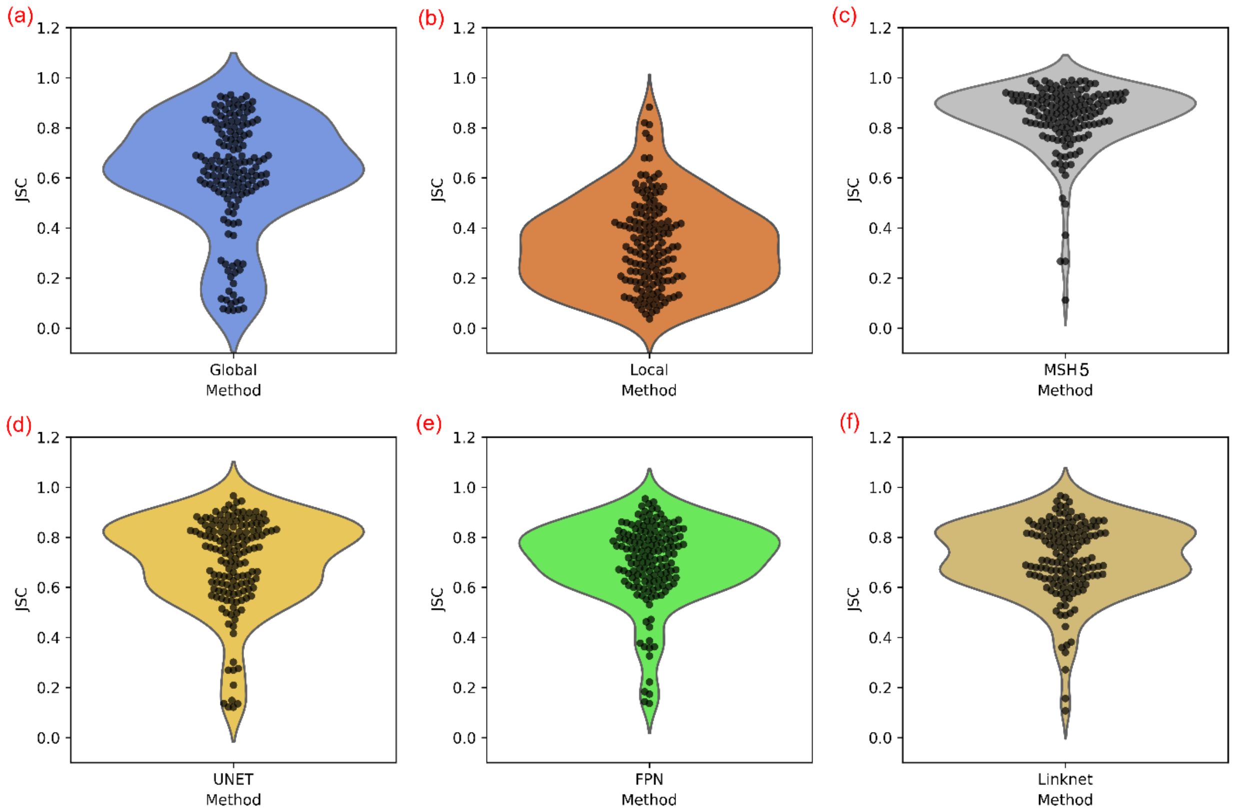

Figure 5. Jaccard Similarity Coefficient distributions for the approaches under consideration. Here, MSH5 denotes the hybrid method.

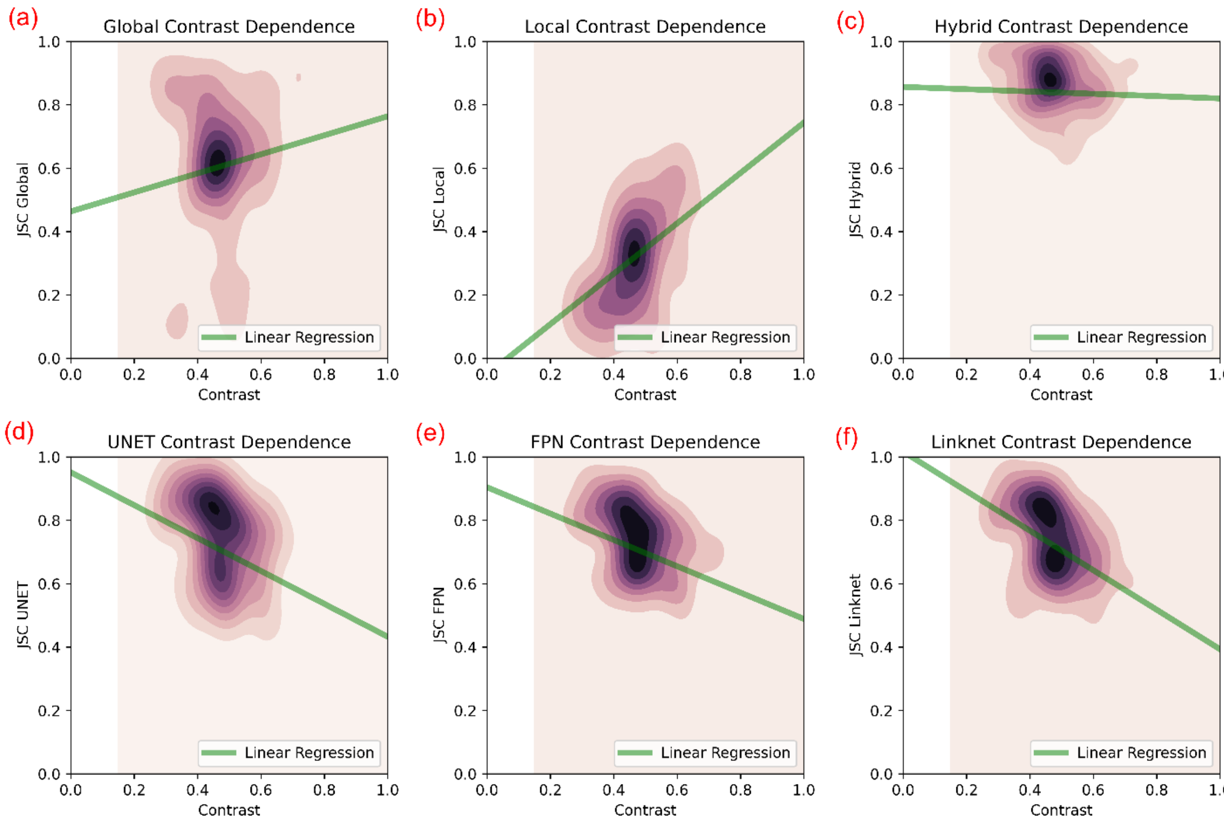

Figure 6. Contrast dependence of the approaches under consideration.

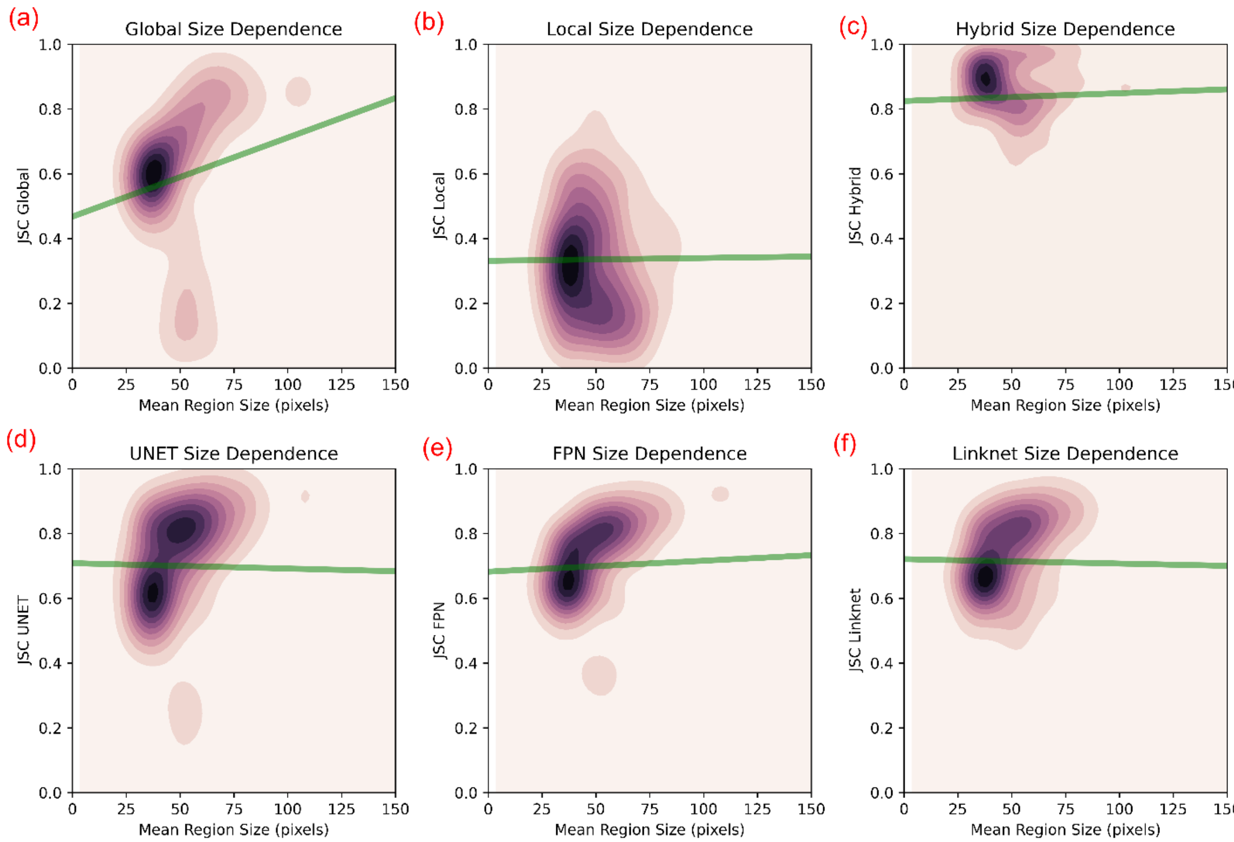

Figure 7. Size dependence of the approaches under consideration.

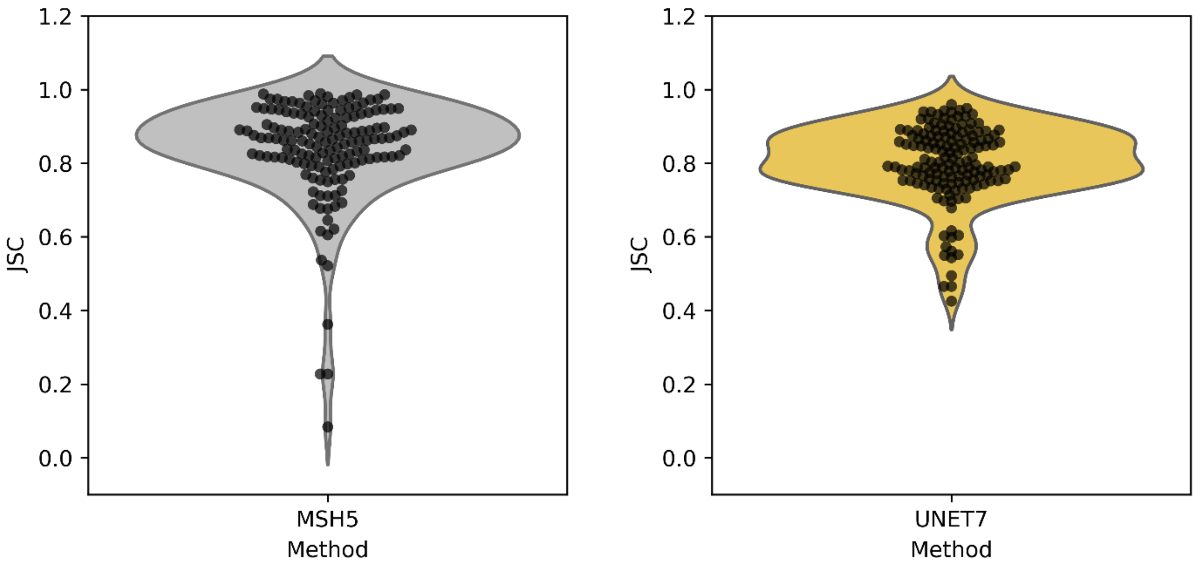

Figure 8. Comparison of the two most accurate models. The hybrid approach (MSH5) still significantly outperforms UNET7, but UNET7 is a significant improvement over approaches shown in Figure 5.

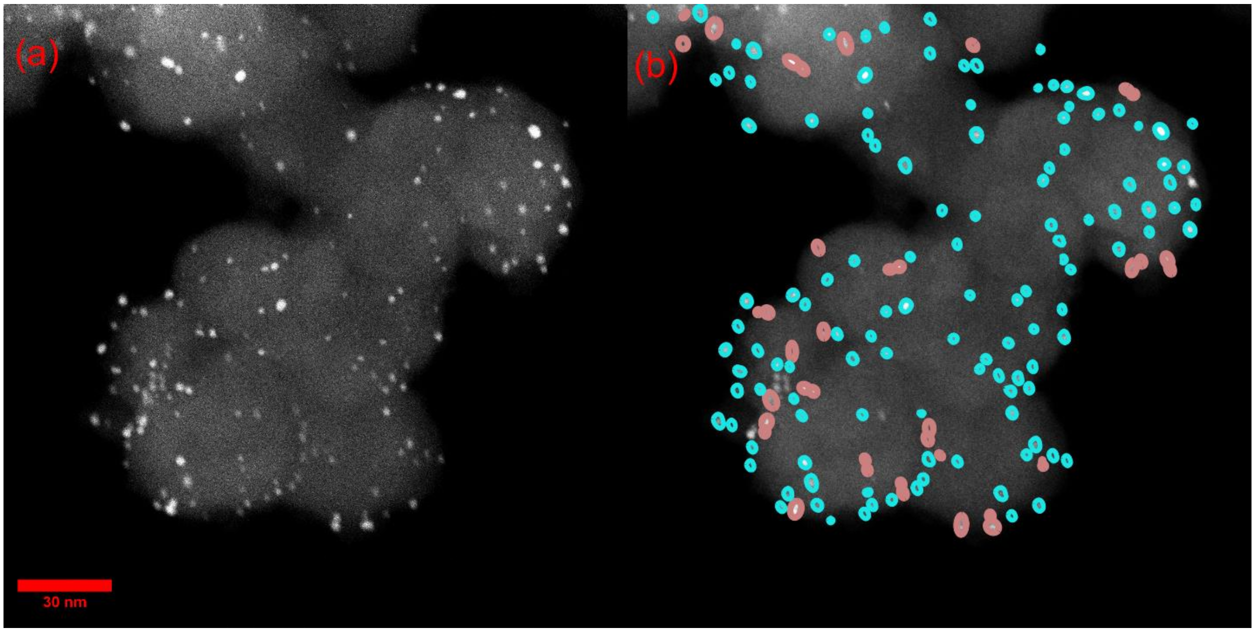

Figure 9. Hybrid thresholding coupled with an instance segmentation paradigm proposed in the literature, applied to a HAADF STEM image.

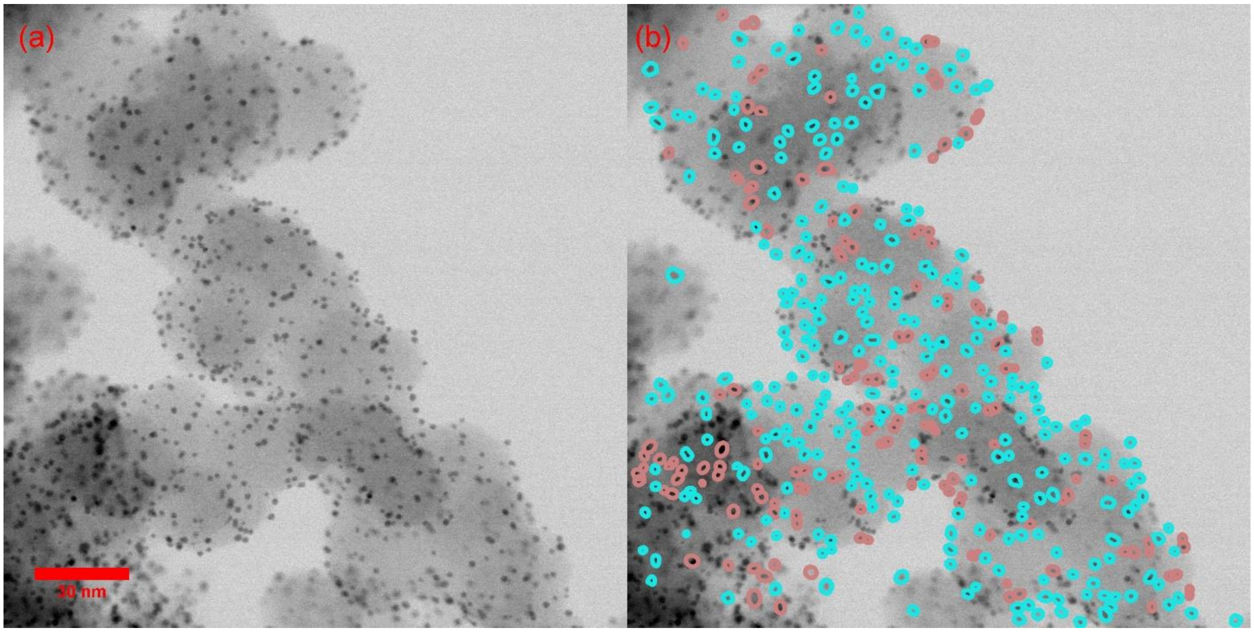

Figure 10. Hybrid thresholding coupled with an instance segmentation paradigm proposed in the literature, applied to a BF STEM image.