On the semantic and instance segmentation of isolated inorganic nanoparticles in TEM-STEM images

Daniel Groom[1]

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

Abstract: Transmission electron microscopy (TEM) has the potential provides many different types of information about inorganic nanoparticles. However, a lack of automated image processing algorithms for segmenting nanoparticles has limited the potential of the instrument to provide statistically meaningful results. In this work, we compare four the performance of four possible methods (global thresholding, local thresholding, hybrid thresholding, and the UNET architecture) for the tasks of semantic and instance segmentation. Broadly speaking, the result of this work is that the UNET architecture performs best on semantic segmentation but worst on instance segmentation, the most likely cause for this discrepancy being undertraining.

1. Introduction

Inorganic nanoparticles promise to provide revolutionary breakthroughs in fields as diverse as medicine and energy. Studying these nanoparticles most often requires use of TEM. In turn, employing TEM to study nanoparticles requires advanced image processing techniques for segmentation. To better understand the need for these segmentation techniques, consider the four TEM images from different imaging modes. In each case, notice how the nanoparticles appear as darker or brighter than the background, although the contrast between the two regions is often visually very low. Moreover, it is also apparent from this figure that TEM images often involve nonuniform backgrounds. In the literature, there are four common approaches to achieve the semantic segmentation of nanoparticles: (1) global thresholding; (2) local thresholding; (3) hybrid thresholding; and (4) employing a deep learning UNET architecture. The first three algorithms are all widely used and available with freeware. However, trained UNET models are in short supply, and none are freely available at the time of writing this manuscript. Figure 2 contains examples of each of the three traditional algorithms used to semantically segment regions associated with nanoparticles for a typical TEM image. Notice how none of the algorithms provide perfect results (though hybrid thresholding comes close). Figure 3 contains the steps that follow semantic segmentation for instance segmentation, which involve separating isolated particles from excluded regions (particles that overlap or are not fully in frame). How do the three publicly available algorithms discussed so far for semantic segmentation fare when submitted to instance segmentation steps? The most relevant way to answer this question is to analyze how well these algorithms reproduce manual particle sizing results. Figure 4 contains such an analysis when each algorithm is applied to the images from Figure 1. From this figure, it is clear that only hybrid thresholding provides reliable results.

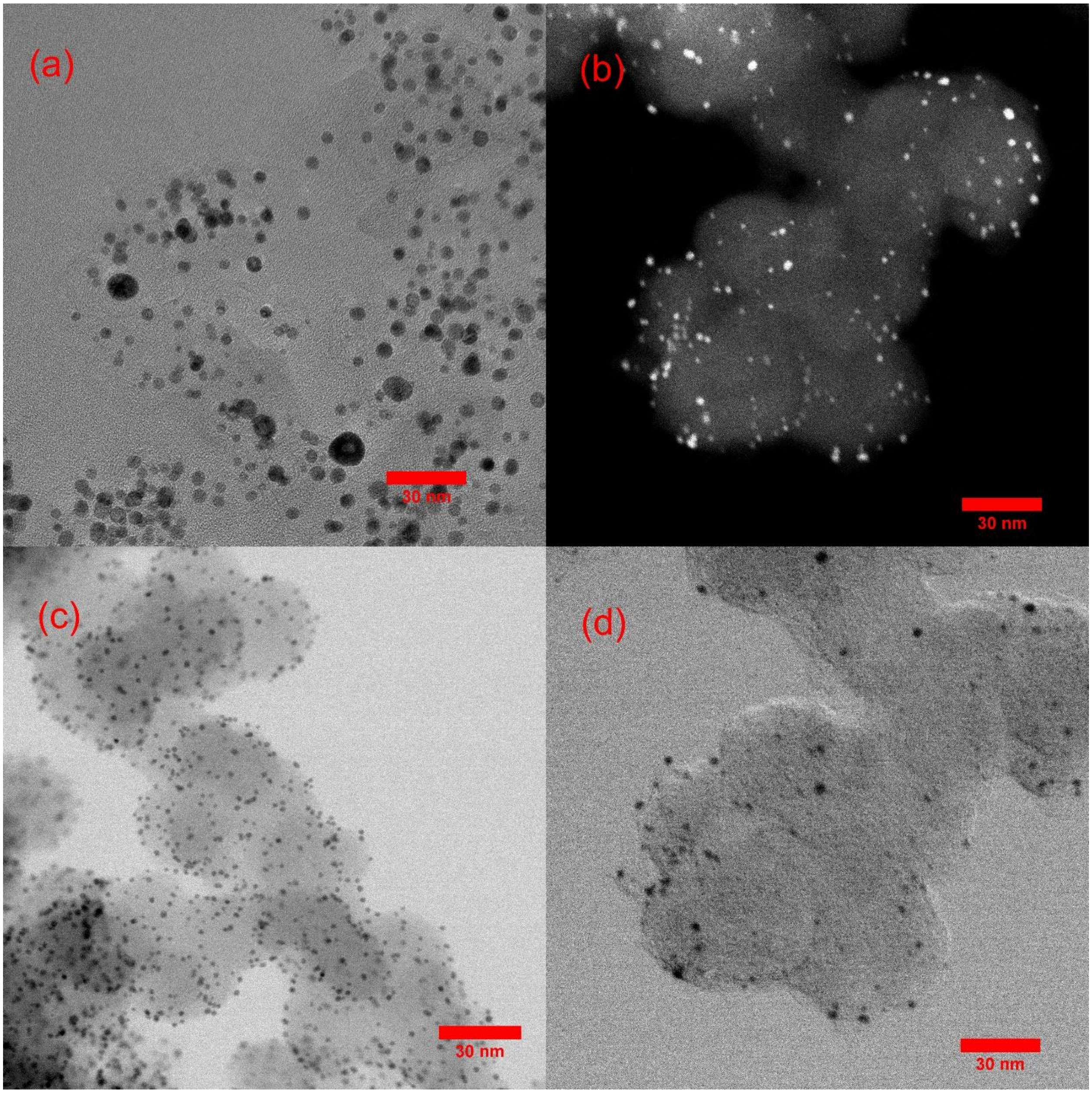

Figure 1. TEM/STEM images using four different modes: (a) Bright-Field TEM (BF TEM); (b) High Angular Annular Dark Field STEM (HAADF STEM); (c) Bright-Field STEM (BF STEM); and (d) Low Angular Annular Dark Field STEM (LAADF STEM).

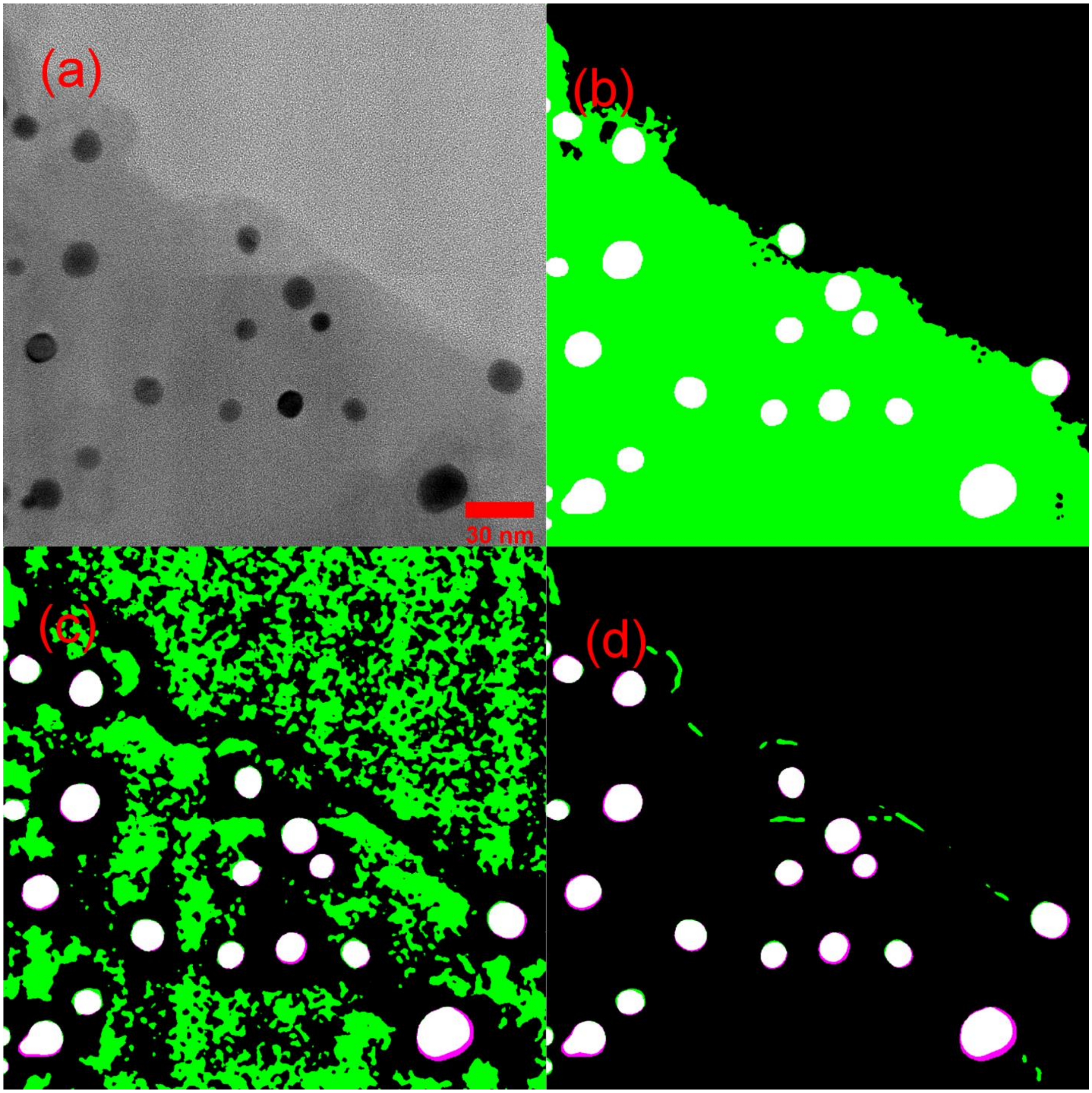

Figure 2. Sematic segmentation results in comparison to ground truth: (a) original BFTEM image; (b) global thresholding; (c) local thresholding; (d) hybrid thresholding. In the images (b) to (d), black and white regions are areas of agreement, whereas colored regions are areas of disagreement.

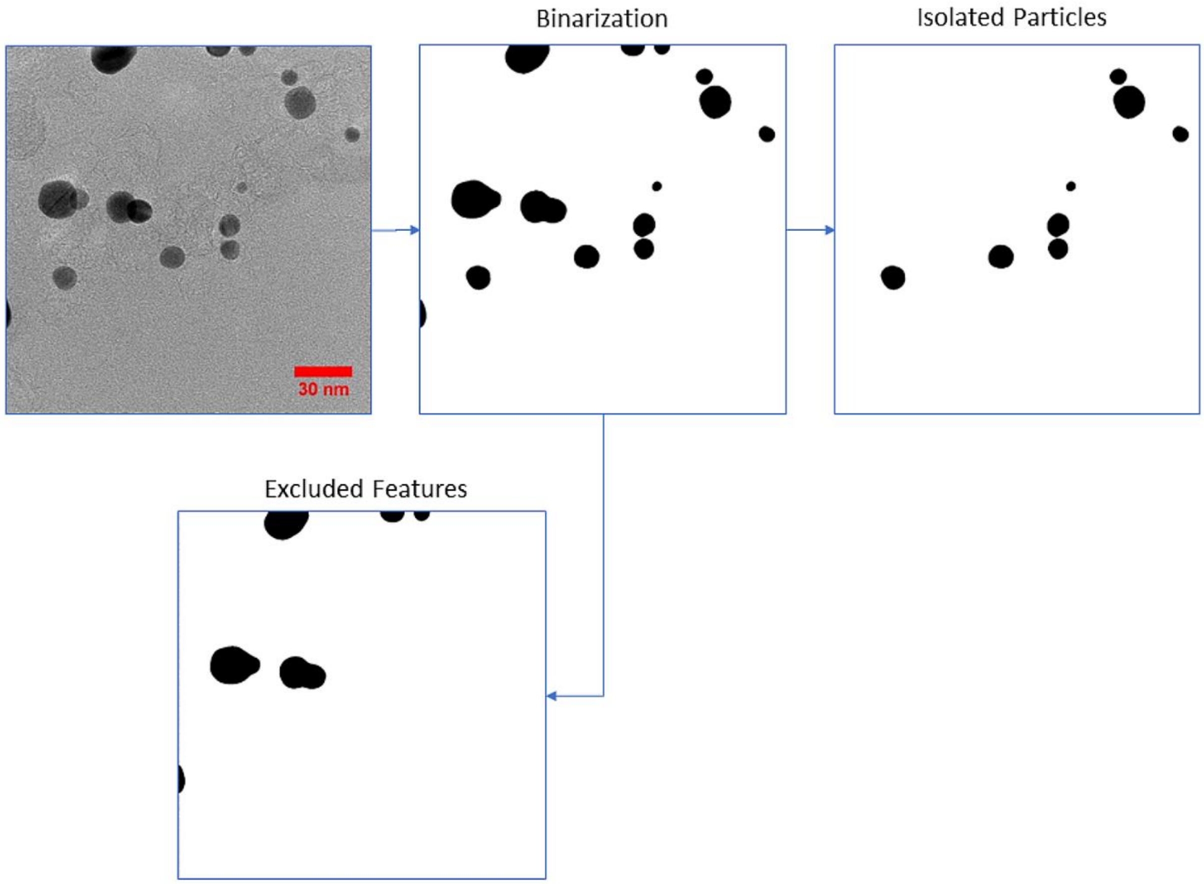

Figure 3. Traditional particle picking paradigm for segmenting isolated particles.

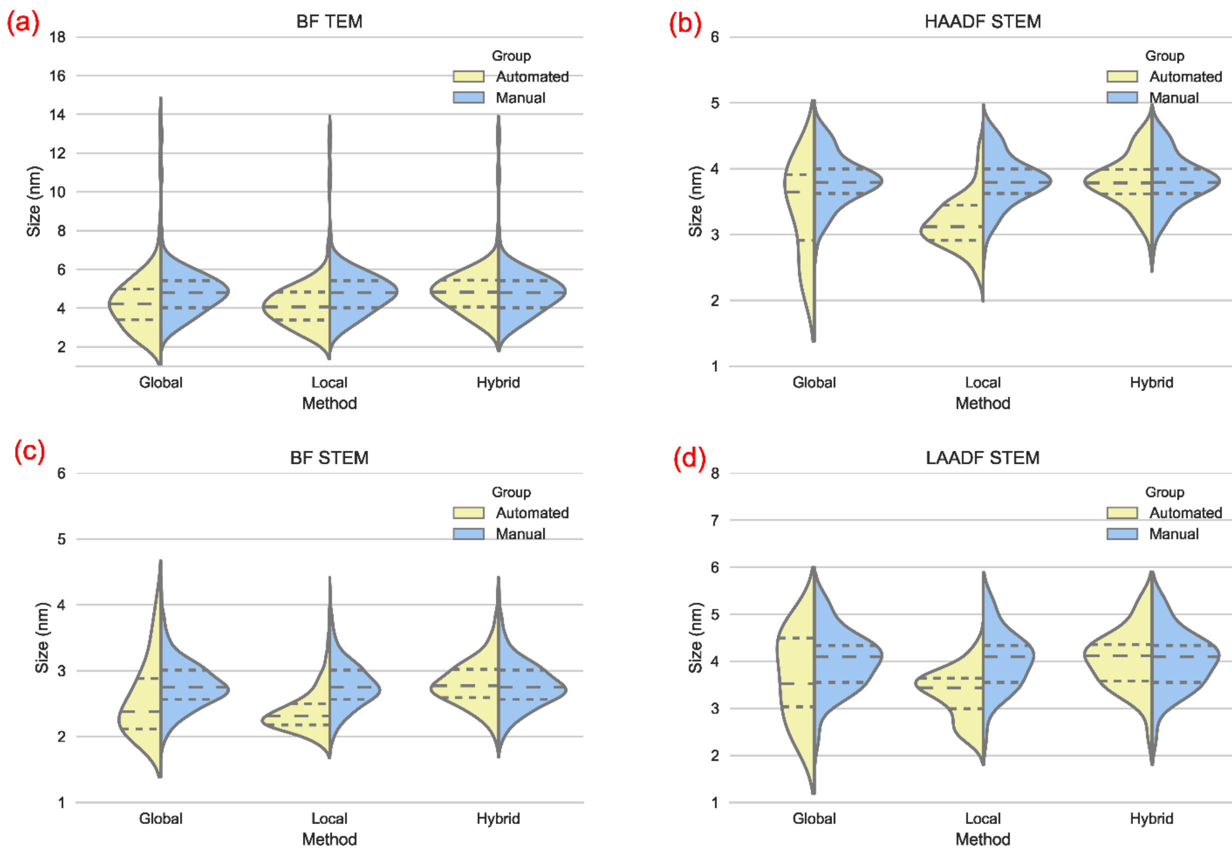

Figure 4. Manual and automated particle sizing results from the images in Figure 1. The four different modalities are (a) BF TEM; (b) HAADF STEM; (c) BF STEM; and (d) LAADF STEM. Automated results for global thresholding, local thresholding, and hybrid thresholding are shown in yellow, whereas manual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

2. Methods

The novel contribution of this manuscript is our unique databases used for training and validation. We begin with 1250 images composed as described in Table 1 below, dividing these images into a training database of 1150 images and a semantic segmentation validation database of 100 images.

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 1250 TEM images

Furthermore, we also have an instance segmentation database of 100 TEM images so that we may study that aspect as well. Our approach is first to train a pretrained UNET model with the database of 1150 images. We then compare the four aforementioned algorithms in their performance on the validation database of 100 images. Finally, we will also study the performance of theses algorithms when it comes to instance segmentation of isolated nanoparticles as described previously.

3. Results and Discussion

Figure 5 contains the results of each algorithm applied to our semantic segmentation database. Here we employ a metric known as the Jaccard Similarity Coefficient (JSC), which ranges from 0 (total failure) to 1 (perfect) based on comparison to ground truth. Notice in (a) and (b) how global thresholding and local thresholding respectively provide poor results. In (c), hybrid thresholding performs reasonably well for a majority of the images, but it does come close to complete failure for a modest amount of the images. Finally, (d) demonstrates that the UNET architecture provides the best results for semantic segmentation, failing in only a small percentage of images.

What about instance segmentation though? Figures 6 and 7 contain the results of our study on 100 images for this issue. Figure 6 contains the JSC distributions for each algorithm as applied to instance segmentation. Notice again how both global thresholding and local thresholding provide unacceptable results. However, hybrid thresholding demonstrates strong reliability for instance segmentation based on this metric. Finally, it is obvious from this figure that the UNET architecture performs worst out of the four algorithms, the most likely reason being undertraining.

Figure 7 quantifies how these JSC observations translate to particle sizing for instance segmentation. In this figure, it is clear that only hybrid thresholding reproduces manual particle sizing, with the other distributions being shifted away from the true distribution by approximately 1nm in each case. This result is particularly surprising given the semantic segmentation performances of each algorithm.

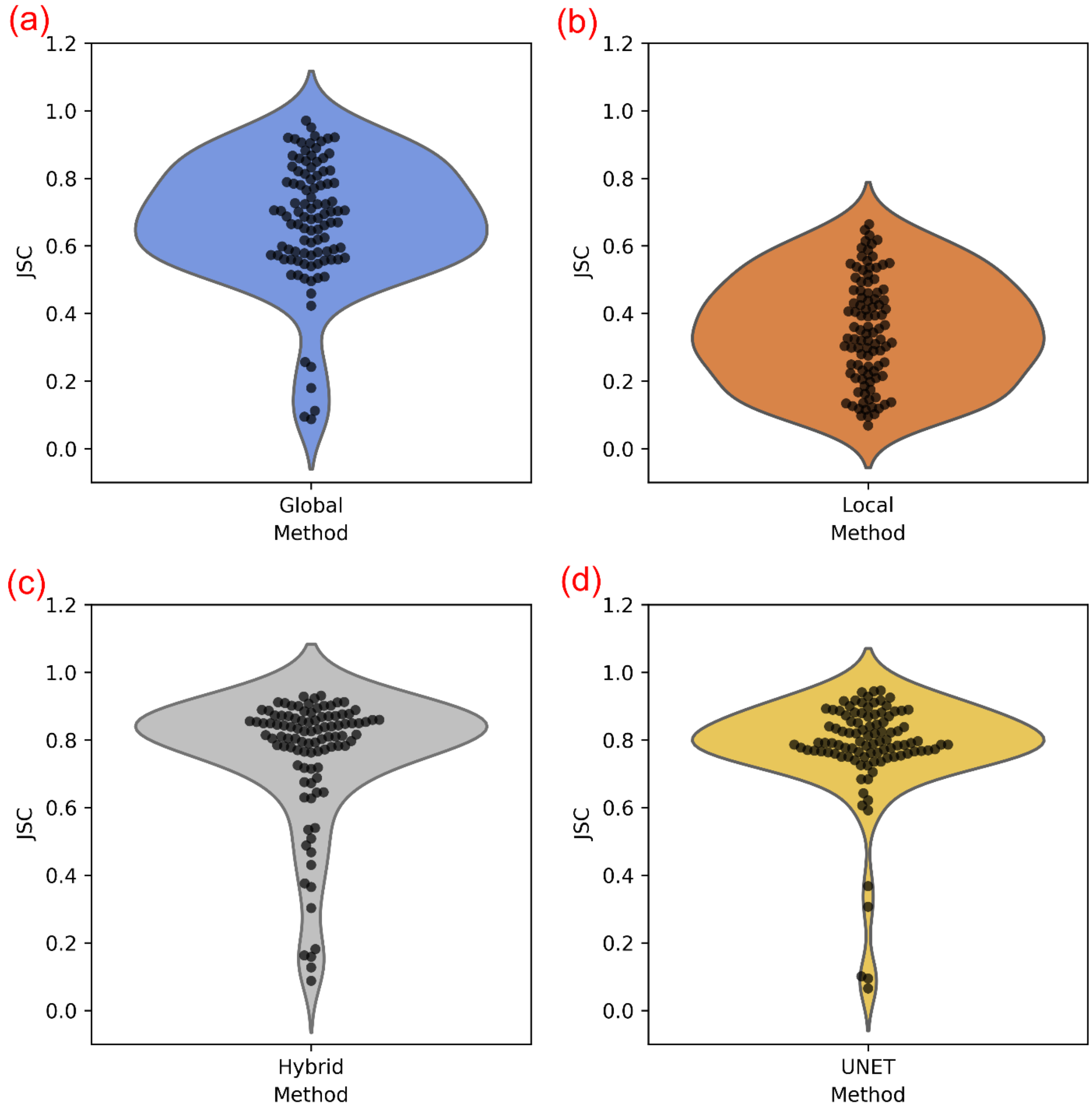

Figure 5. JSC distributions for semantic segmentation algorithms: (a) global thresholding; (b) local thresholding; (c) hybrid thresholding; and (d) the UNET architecture.

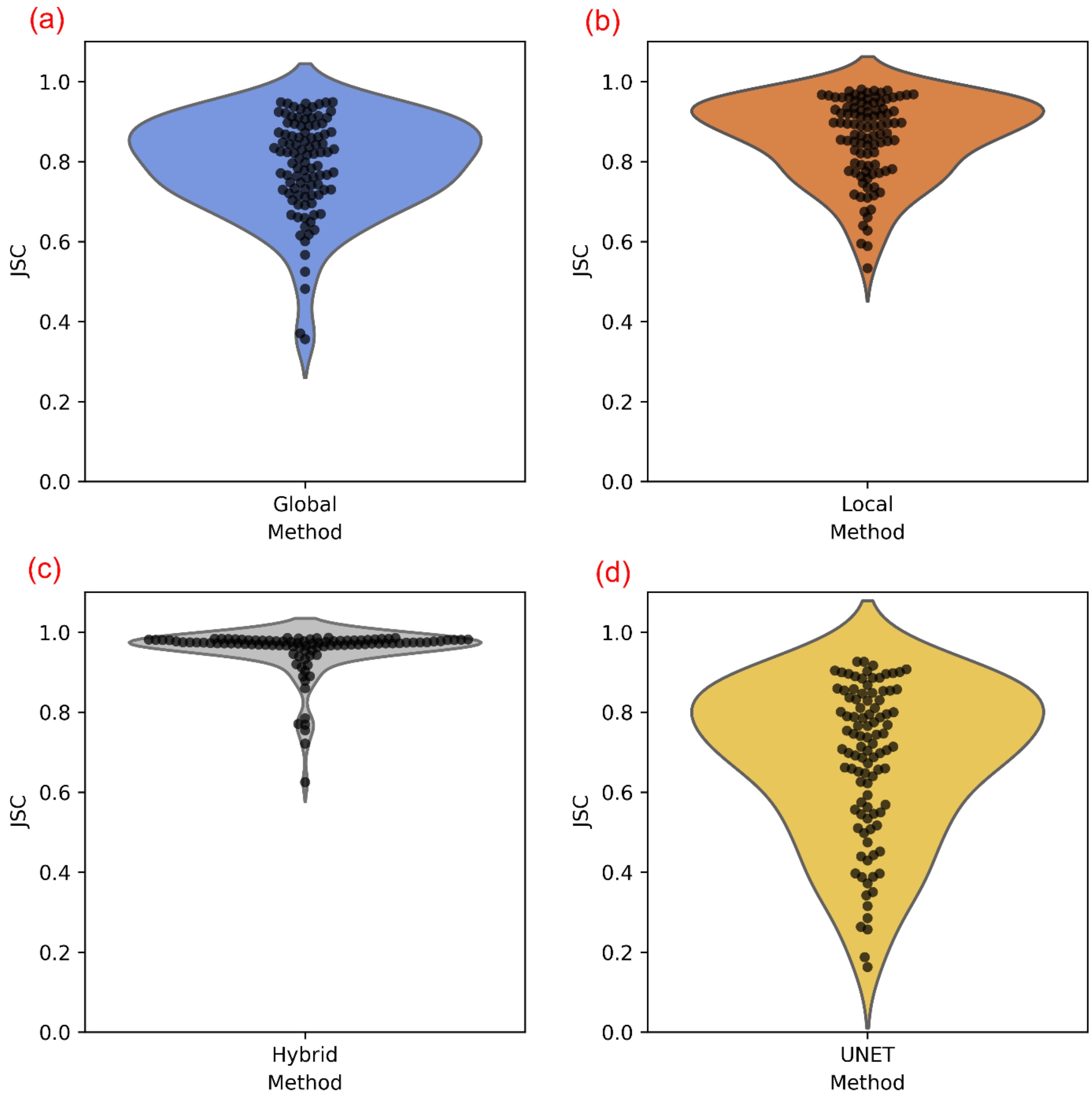

Figure 6. JSC distributions for instance segmentation algorithms (isolated particles only): (a) global thresholding; (b) local thresholding; (c) hybrid thresholding; and (d) the UNET architecture.

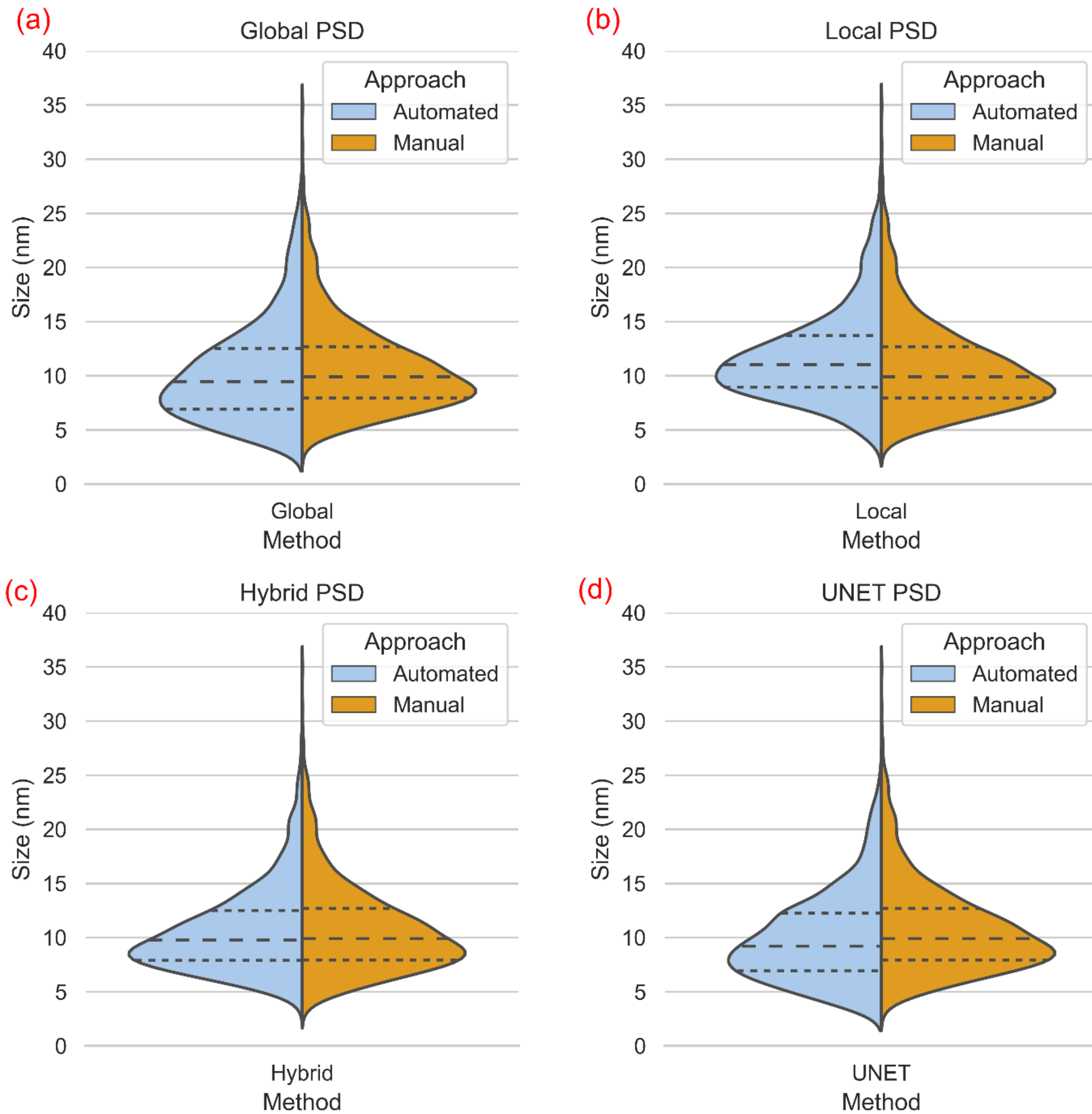

Figure 7. Comparison of automated to manual PSDs for (a) global thresholding, (b) local thresholding, (c) hybrid thresholding, and (d) the UNET architecture.

4. Conclusion

This paper demonstrates that global thresholding and local thresholding provide poor results for both semantic and instance segmentation on our databases. Hybrid thresholding provides somewhat mixed results in that it performs relatively poorly with semantic segmentation but very well with instance segmentation. Finally, the UNET architecture also offers mixed performance in that it performs very well with semantic segmentation but very poorly with instance segmentation. We blame undertraining on this final result and propose that future work focus on training a pretrained network with a database of 5,000 TEM images.