

<https://doi.org/10.1038/s41524-024-01302-w>

Uncertainty-aware particle segmentation for electron microscopy at varied length scales

Check for updates

Luca Rettenberger¹, Nathan J. Szymanski², Yan Zeng³, Jan Schuetzke¹, Shilong Wang², Gerbrand Ceder² & Markus Reischl¹ ✉

Electron microscopy is indispensable for examining the morphology and composition of solid materials at the sub-micron scale. To study the powder samples that are widely used in materials development, scanning electron microscopes (SEMs) are increasingly used at the laboratory scale to generate large datasets with hundreds of images. Parsing these images to identify distinct particles and determine their morphology requires careful analysis, and automating this process remains challenging. In this work, we enhance the Mask R-CNN architecture to develop a method for automated segmentation of particles in SEM images. We address several challenges inherent to measurements, such as image blur and particle agglomeration. Moreover, our method accounts for prediction uncertainty when such issues prevent accurate segmentation of a particle. Recognizing that disparate length scales are often present in large datasets, we use this framework to create two models that are separately trained to handle images obtained at low or high magnification. By testing these models on a variety of inorganic samples, our approach to particle segmentation surpasses an established automated segmentation method and yields comparable results to the predictions of three domain experts, revealing comparable accuracy while requiring a fraction of the time. These findings highlight the potential of deep learning in advancing autonomous workflows for materials characterization.

Electron microscopy offers detailed insight into the structure, composition, and morphology of inorganic materials. This technique is widely used to characterize powder samples, which are prevalent at the laboratory scale as well as in a variety of applications such as energy storage and ceramics. Individual particles within a powder often display a wide range of shapes and sizes, which can have a significant influence on the macroscopic properties of the corresponding material^{1,2}. For the evaluation of particle morphology in powder samples, desktop Scanning Electron Microscopes (SEMs) are becoming commonplace in laboratories and industrial settings, allowing images to be produced at an unprecedented rate^{3,4}. To keep pace with this surge in data, new and improved methods are needed to automate the analysis of SEM images and produce meaningful conclusions that can aid in materials design.

A key step in analyzing SEM images is the identification of distinct particles, otherwise known as particle segmentation. At present, segmentation is often performed manually or with traditional methods such as thresholding

or edge detection⁵⁻⁷. Previous efforts have created hand-crafted kernels in the form of convolution matrices, which are capable of detecting individual particles in SEM images⁸. More generally, convolution matrices for particle segmentation can be optimized through Deep Learning (DL), which has gained traction in recent years. For segmentation in particular, Convolutional Neural Networks (CNNs) based on the U-Net architecture have been most widely used⁹. This architecture compresses the input image into a compact feature space and then symmetrically expands it to generate the desired segmentation mask. The key innovation of U-Net is the inclusion of skip connections, enabling distant layers to share information and improve feature learning and localization. Although initially designed for biological applications, U-Net models have been extended to a variety of applications in electron microscopy. These include vacancy and dopant detection in Transmission Electron Microscopy (TEM)¹⁰, as well as particle detection in SEM^{11,12}. Such models have been reported to provide state-of-the-art performance when compared with more traditional methods for segmentation¹³⁻¹⁵.

¹Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. ²Department of Materials Science & Engineering UC Berkeley, Berkeley, CA 94720, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

✉ e-mail: markus.reischl@kit.edu

Despite the success of U-Net, recent studies reveal its limitations in complex datasets with overlapping instances, showing inferior performance compared to Mask R-CNN, a more resilient and modern CNN architecture¹⁶. Further, some early work has been reported on using Mask R-CNN to solve problems in materials science and chemistry, with multiple reviews suggesting Mask R-CNN over U-Net^{17–19}. For example, there have been several models trained to segment particles^{20–22}, nanowires²³, and cavities^{24,25} in electron microscopy with an accuracy that exceeds more traditional approaches.

When analyzing scans of powder samples using desktop SEMs, both manual analysis and DL-based approaches face similar challenges due to limited resolutions and complex morphology. Challenges such as particle agglomeration, suspension stability, and the effects of sonication complicate measurements²⁶. Additionally, addressing the difficulties in obtaining accurate 2D scans from three-dimensional particles further adds to the complexity of nanoparticle analysis with SEMs. Hence, the images frequently display a high degree of blur, and the particles often form agglomerates where many instances overlap. Developing a general segmentation model that can handle these challenging cases requires new training data obtained from a variety of samples that encompass each of the aforementioned issues, as such datasets do not exist yet²⁷. Further, in cases where these issues preclude accurate segmentation, a measure of prediction uncertainty would be beneficial to avoid making over-confident predictions that lead to incorrect conclusions regarding the data. Novel approaches could also be further enhanced by considering the significant uncertainty within material discovery. Further, extensive validation by human experts remains pivotal in ensuring practical applicability. Additionally, it remains essential for methods not only to align with human expertise but also to demonstrate universality across diverse chemical compounds.

In this work, we introduce an approach for segmenting particles in desktop SEM images by enhancing the popular Mask R-CNN architecture. To train and validate our models, we collected 90 experimental SEM images from a variety of inorganic powder samples, each hand-labeled by domain experts to outline individual particles. These images are split into two separate datasets, corresponding to SEM measurements performed at low or high magnification. Each dataset is used to train our enhanced segmentation model based on Mask R-CNN, augmented to estimate uncertainty, allowing the trained models to output a measure of confidence associated with each segmented particle. On a holdout set of images reserved for testing, we demonstrate that our uncertainty-aware models outperform segmentation methods based on the U-Net architecture. To prove plausibility and applicability, we test our models on a second, much larger, set of 288 SEM images obtained from LiCoO₂ powder, a composition the models have not encountered during training. This data is also labeled by three domain experts and we show that our Mask R-CNN models produce results comparable to those ground-truth annotations. With this, we show that our approach can transfer to previously unseen data. Notably, our models completed this task in just three minutes on a desktop processor, while domain experts required an average of 265 minutes for their labeling. These findings showcase the benefits of automated segmentation and suggest that DL models are well suited for integration with high-throughput and closed-loop workflows. All code and data discussed here are openly accessible, including our uncertainty-aware models and datasets: <https://github.com/lrettenberger/Uncertainty-Aware-Particle-Segmentation-for-SEM>. We encourage the community to utilize this repository as a foundation for further development and exploration.

Results

Hand-labeled training data

Using a Phenom Desktop SEM from Thermo Fisher Scientific, we collect 90 images from 10 different samples, each containing one of the following compounds: NaAlSiO₄, Cu₃(PO₄)₂, MgO, Mn₃O₄, Na₂CO₃, TiO₂, BaCO₃, SiO₂, CaTiO₃, and BaCuO₂. These images are classified into two distinct

categories: one corresponds to images acquired at magnifications below 10,000× (denoted as low magnification), while the other contains images acquired at magnifications over 10,000× (denoted as high magnification). We acquire 50 images at low magnification, each with a resolution of 1920 × 1200 pixels, and 40 images at high magnification with a resolution of 7680 × 4800 pixels. During image capturing, additional images were added iteratively until the DL model did not significantly improve in accuracy with the inclusion of new samples. Consequently, fewer images were collected in the high-magnification case. To avoid introducing an implicit bias into the data, all images are resized to a uniform resolution of 1920 × 1200 pixels before being fed into the DL model. From preliminary tests (Supplementary Table 5), we found that performance was improved by separately training models on the data collected at low or high magnification, as these two regimes tend to produce visibly distinct images – for instance, with varied amounts of blur.

All SEM images are hand-labeled by domain experts, who are assigned the task of segmenting distinct particles within each sample. Given the challenging nature of labeling desktop SEM images, particularly when dealing with overlapping particles, the experts provide each segmented particle with one of two label types. *Certain* labels are assigned to particles whose boundaries can be drawn with a high degree of subjective confidence, while *uncertain* labels are given to particles with less well-defined boundaries. Several examples of these labels are shown in Fig. 1. After labeling, the images are prepared for training and validation as detailed in Supplementary Table 1 and Supplementary Figure 1.

Uncertainty-aware Mask R-CNN

For the segmentation of particles in SEM images, we consider two possible architectures: U-Net and Mask R-CNN. The first method features a U-shaped architecture with an encoder for feature extraction and a decoder for creating segmentation masks. It generally performs well in semantic segmentation, where pixels are assigned to predefined categories. However, it is not inherently designed for instance segmentation, where each individual object within a category must be separated. As such, it can struggle with overlapping objects which are highly prevalent in SEM data. In contrast, the Mask R-CNN architecture was developed specifically for instance segmentation. It operates by initially extracting relevant features from the provided image. These features are then used to identify regions of interest, from which bounding boxes are created and segmentation masks are refined to segment each individual object. This enables Mask R-CNN to perform well on images with many overlapping instances, motivating our choice to use this architecture for SEM analysis. Figure 2 illustrates the architecture of the segmentation models employed in this work.

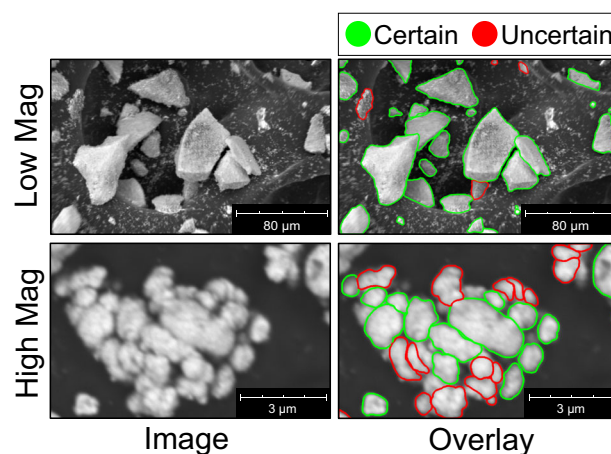


Fig. 1 | Two examples of SEM images taken from our hand-labeled datasets acquired at low (top) and high magnification (bottom). In the right panels, colored curves represent particle boundaries outlined by domain experts, with green indicating certain labels and red denoting uncertain labels.

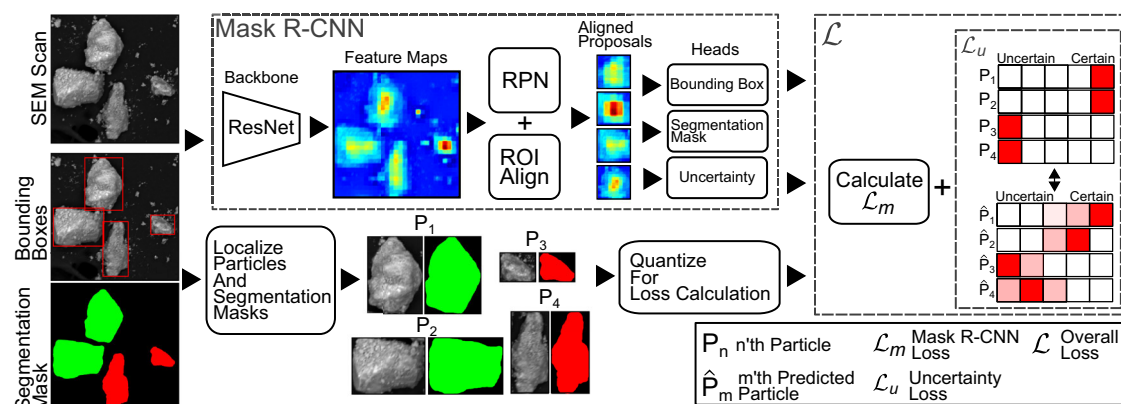


Fig. 2 | The Mask R-CNN architecture uses a ResNet-50 backbone to extract feature maps from provided SEM images. A Region Proposal Network (RPN) generates Regions of Interest (RoIs) from these features, which are then aligned to a consistent size using ROI align. The Mask R-CNN heads (bounding box and segmentation mask) as well as our uncertainty head process the aligned proposals to

generate predictions \hat{P} . The bounding boxes and segmentation masks are transformed into numerical ground-truth values P . The Mask R-CNN loss \mathcal{L}_m and uncertainty loss \mathcal{L}_u are calculated and combined to obtain the overall loss \mathcal{L} . For \mathcal{L}_u , the uncertainty head outputs \hat{P} as a distribution with multiple bins.

The Mask R-CNN framework takes SEM images as input and processes them using a ResNet-based neural network, referred to as the backbone, that transforms the images into feature maps, which are visual representations of the input designed to capture its most significant patterns. Region of Interests (RoIs) are generated from these feature maps, corresponding to parts of the image suspected to contain particles. Because particles often vary in size, each ROI can have unique dimensions. To facilitate uniform processing, ROI align transforms the ROIs to a consistent size. These resized ROIs are then passed to a set of neural networks referred to as heads each reserved for a single task. In the context of Mask R-CNN, the bounding box regression head refines the ROIs to focus on specific particle instances, while the mask head segments each particle by generating pixel-wise masks.

In addition to the well-established heads, we introduce an uncertainty head that generates a confidence score for each segmented particle. A higher confidence score is designed to signal more accurate predictions, while particles with lower confidence should be taken with caution. These scores are calibrated based on the segmentation masks and labels (certain or uncertain) provided by domain experts on our hand-labeled set of SEM images. Both the loss from the Mask R-CNN and our uncertainty head are minimized during training, as outlined in Fig. 2. Further details on the training process are also provided in the Methods section.

Segmentation results

We partition the hand-labeled SEM images into three distinct sets reserved for training (65% of the data), validation (15%), and testing (20%) purposes. To gauge the accuracy of each model, we employ the Aggregated Jaccard Index (AJI^+), a robust metric for instance segmentation that quantifies the overlap between ground truth objects and their corresponding predictions, considering both localization accuracy and segmentation quality simultaneously²⁸. It does so by measuring the intersections and unions between ground truth and predicted segmentation masks of matching pairs. The AJI^+ is then computed as the ratio of the total intersections to the total unions, where unmatched objects are also accounted for by incorporating them into the union count. A higher AJI^+ value (within the range of 0 to 1) therefore signifies more accurate and precise predictions of segmented areas.

To demonstrate the superior performance of our Mask R-CNN version compared to conventional methods, we compare it to a widely used U-Net-based approach²⁹. Details regarding the configurations of these models are provided in Supplementary Table 2 and Supplementary Table 3. Both the U-Net and Mask R-CNN models are trained and validated on the same images from our hand-labeled datasets. Final results are generated from the test dataset, from which the resulting AJI^+ values are plotted in Fig. 3. These

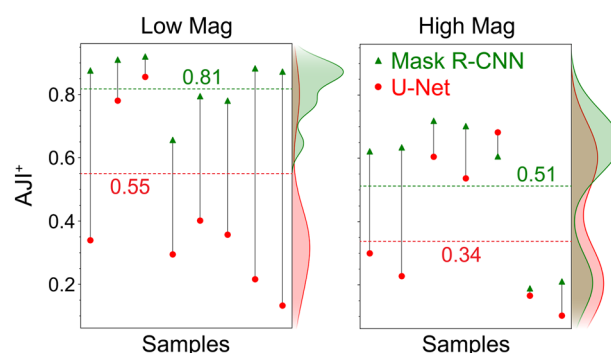
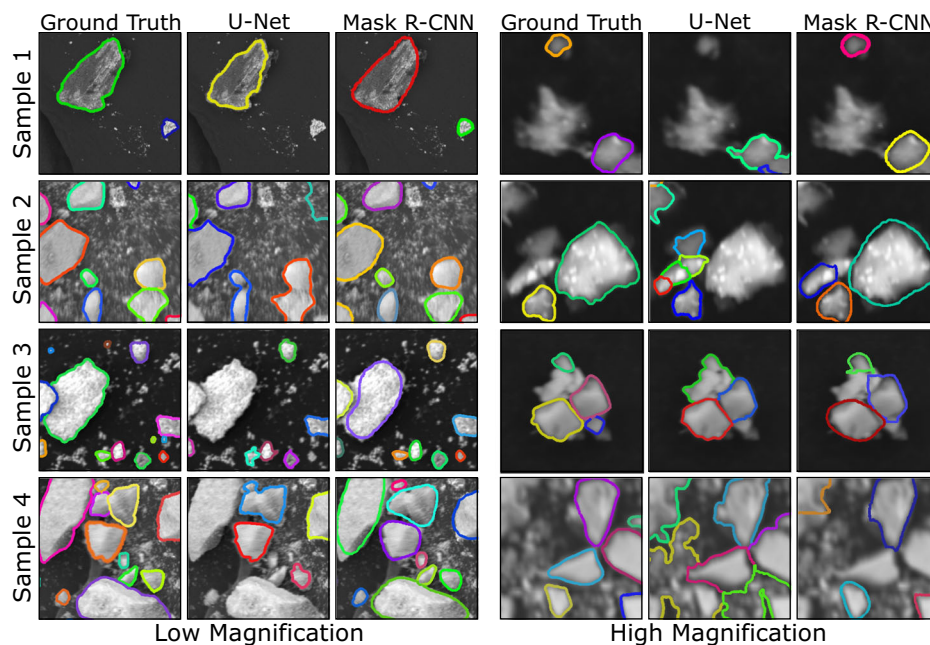


Fig. 3 | The AJI^+ metric produced by segmentation models when applied to SEM images acquired at low (left panel) and high magnification (right panel). Green triangles correspond to results from Mask R-CNN models, while red circles are from U-Net models. Results from the same image are linked with a black line. A density representation of the AJI^+ values is shown on the right axis of each panel. Horizontal dashed lines represent average AJI^+ scores from each method. Mask R-CNN is superior to U-Net in all but one sample.

plots reveal that our Mask R-CNN models outperform U-Net on 14 out of the 15 images that are tested. When applied to SEM images that are captured at low magnification, the Mask R-CNN model yields an average AJI^+ score of 0.81, whereas U-Net provides a much lower score of 0.55. A similar performance gap is found on the images acquired at high magnification, with the Mask R-CNN model achieving a moderate AJI^+ score of 0.51 as compared to only 0.34 from U-Net.

To better understand why the Mask R-CNN models outperform U-Net, we visualize in Fig. 4 the segmented areas of four samples from each dataset used for testing. In the first sample acquired at low magnification, there is a clear particle that is accurately segmented by Mask R-CNN but missed by U-Net, possibly due to its small size and irregular shape. U-Net also performs poorly on the second sample, where it appears to detect particles based on a brightness threshold, without recognizing their actual structure. This leads to an incorrect grouping of small particles which should be separate but are mistakenly segmented as one large particle. U-Net also produces a spurious segmentation of the background in the top right of the image, where no particles are present. In contrast, Mask R-CNN correctly segments all the individual particles in this sample, even separating those clustered together. In the third sample acquired at low magnification, U-Net again struggles with particle overlap. It fails to segment two of the largest particles, whereas Mask R-CNN accurately identifies their boundaries.

Fig. 4 | Four samples taken from the SEM datasets obtained at low and high magnification. The ground truth represents overlays that are manually created by domain experts. Overlays predicted by U-Net and Mask R-CNN models are also shown. In each image, the overlay colors are chosen arbitrarily.



However, both models fail to detect four of the smaller particles in this image. Similar effects are observed in the fourth sample, where U-Net fails to segment two large particles while also incorrectly grouping smaller particles that overlap but should be separate. Mask R-CNN offers improved segmentation of the particles, though it still incorrectly combines two of the particles that have substantial overlap. Overall, Mask R-CNN demonstrates superior performance in segmenting diverse shapes and sizes. Its precision in distinguishing closely clustered particles showcases its ability to differentiate complex structures. Moreover, Mask R-CNN's proficiency in recognition of individual objects and differentiation from the background signifies its potential for enhanced segmentation, even in challenging scenarios where other methods encounter difficulties.

The performance gap between U-Net and Mask R-CNN becomes even more pronounced when these models are applied to images acquired at high magnification, which tend to have notably increased blur. U-Net inaccurately segments two particles with imprecise borders in the lower right corner of the first sample, while Mask R-CNN outlines borders that agree well with the labels provided by domain experts. The second sample produces similar results, whereby U-Net incorrectly segments four small particles while missing a much larger one that should be segmented. Mask R-CNN provides more accurate segmentation masks in this sample, though it does generate one false positive surrounding a bright area of the image that one may suspect to be a particle. In the final two samples, we see that U-Net incorrectly groups many smaller particles together into one segmented area, similar to what was observed in the images acquired at low magnification. The predictions from U-Net are particularly inaccurate on the fourth sample, where the model segments very large areas of the image without detecting many of the individual particles that exist. This shortcoming can be attributed to the large amount of blur and low contrast in this image, which causes the particles to blend in with the background. Mask R-CNN successfully avoids segmenting the background in this sample and correctly identifies several of the larger particles, though it does miss two smaller particles. Even with difficult high-magnification images, Mask R-CNN excels at detecting individual particles. Despite occasional false positives, it outperforms U-Net by effectively isolating objects in blurry, low-contrast backgrounds. The occasional disagreements between ground truth and Mask R-CNN mainly stem from the inherent labeling uncertainty due to limited resolutions and complex morphology. Mask R-CNN mostly identifies particle-like regions reliably, while U-Net often misclassifies noise. This underscores Mask R-CNN's strength in adapting to tough image

conditions, positioning it as a promising framework for precise object delineation in complex visual environments.

Uncertainty estimation

In this section, we examine the Mask R-CNN model's ability to assess its own confidence in the segmentation masks it provides for each particle. Figure 5a shows the distribution of prediction confidence generated by our models when applied to the hand-labeled SEM images reserved for testing. These results are categorized based on whether the segmented particles were labeled as *certain* or *uncertain* by domain experts. We observe a clear correlation between the prediction confidence of a particle and its associated label. At low magnification, nearly all the particles labeled as certain by domain experts are segmented with a confidence higher than 50% by the Mask R-CNN model. For the particles labeled as uncertain, the majority of predictions fall below 50%—though some do exist with higher confidence. A similar trend is observed on the images acquired at high magnification, where most of the particles labeled as certain are segmented with a confidence exceeding 50%. We note that in this case, there are many more particles labeled as uncertain by the domain experts, likely due to the large amount of blur that is present at high magnification. The Mask R-CNN accurately accounts for this effect by providing segmentation masks with much lower confidence (often $\leq 50\%$).

To help visualize the role of uncertainty in particle segmentation, we present in Fig. 5b two examples from our test set. These examples show particles labeled as certain (green) or uncertain (red) by domain experts. For comparison, predictions from the Mask R-CNN model are also shown, where segmentation masks with a confidence $> 50\%$ are colored green and those with a confidence $\leq 50\%$ are colored red. At low magnification, we see that prediction confidence is correlated with particle size, as the large particle is segmented with a higher confidence than the small one, matching the expert's labels. However, it is not always the case that particle size influences a label's confidence. For the image acquired at high magnification, the largest particle is segmented with a lower confidence than two smaller ones, again matching the labels provided by domain experts. In this case, it appears that the prediction confidence is more affected by how well defined the particle boundary is, regardless of its size. These results confirm that our models can effectively distinguish between particles that should be segmented with low or high certainty in a variety of images with varied quality and particle morphology. Additional analyses displaying the prediction accuracy split by certainty and evaluating the correlation between particle

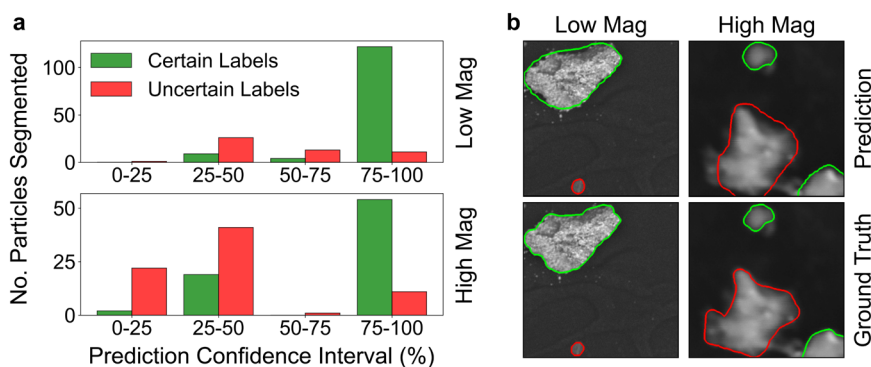


Fig. 5 | Confidence distribution and segmentation accuracy of Mask R-CNN Models. **a** Shows the distribution of prediction confidence generated by the Mask R-CNN models when applied to particle segmentation on images obtained at low (top panel) and high (bottom panel) magnification. The green bars correspond to particles that were labeled as “Certain” by domain experts, while the red bars correspond to particles labeled “Uncertain.” **b** Displays two sample images for low and

high magnification that were segmented by the Mask R-CNN models are shown in the top panels. For comparison, the labels provided by domain experts are shown in the bottom panels. Green and red curves represent certain and uncertain labels, respectively. Predictions from the Mask R-CNN are considered uncertain when their confidence is $\leq 50\%$.

size and predicted confidence can be found in Supplementary Table 6 and Supplementary Fig. 4.

Analysis of particle size

We apply our Mask R-CNN model to characterize the particle size distribution of a powder sample. This is done for a sample of LiCoO_2 , which is commonly used as a cathode in modern batteries. The powder was purchased from Sigma Aldrich. After dispensing the powder onto a sample stub for SEM, the Phenom XL desktop SEM is used to acquire 288 images spanning a uniform 12×24 grid, with each individual image covering an 80×80 m region of the sample. These images are processed by the Mask R-CNN model, which is trained at low magnification, allowing us to evaluate its accuracy in segmenting particles. Further, as this type of powder sample was not used during the training of the model, this also evaluates the transferability to arbitrary inorganic particle detections. The size distribution of the detected particles is computed from the segmented areas of each image, including both certain and uncertain predictions. For comparison, we task three separate domain experts with labeling the same set of images. They also recorded how much time is required to complete this task.

Our Mask R-CNN model predicts an average particle area of 37 m, corresponding to an average diameter of approximately 6.7 m when assuming spherical particles, which agrees well with the expected particle size of typical LiCoO_2 powders (ranging from 5–10 m). In comparison, the labels provided by the domain experts correspond to an average particle area of 31 m, closely matching the prediction of Mask R-CNN. The particle size distribution predicted by our Mask R-CNN model is shown in Fig. 6a, and comparable plots representing the labels made by three domain experts are provided in Fig. 6b–d. The key characteristics of the distributions from Mask R-CNN and the domain experts appear qualitatively similar, with a peak in the number of particles that have an area of about 10 m and a long tail of larger particles with a size reaching 200 m.

The sizes of particles identified by the three experts range from 0.01% to 9.41% of the total image, while Mask R-CNN predicts sizes ranging from 0.08% to 9.41%. Although Mask R-CNN can detect particles as small as 0.08% of the total image, the smallest particle annotated by the experts (0.01%) is even smaller. Hence, detecting very small particles challenges Mask R-CNN. However, our high-magnification model is poised to address this, ensuring thorough detection and analysis across all scales.

Interestingly, there is noticeable disagreement among the experts themselves, whose labels produce an average particle size ranging from 21 to 45 m. Using the mean AJI⁺ across all samples among the experts and our model to calculate the agreements further supports this observation. The agreement between expert 1 and expert 2 calculates to 64.61 ± 19.5 , between

expert 1 and expert 3 to 57.93 ± 22.5 , and between expert 2 and expert 3 to 56.32 ± 21.3 . The agreement between Mask R-CNN and expert 1 is 61.00 ± 21.3 , with expert 2 64.12 ± 20.9 , and with expert 3 60.31 ± 22.7 . This displays that the predictions made by the mask R-CNN model generally agree well with labels created by the experts on average, despite the significant variations that exist among them.

This showcases the uncertainty that is prevalent in particle segmentation, which becomes increasingly difficult when dealing with small particle size. Indeed, Experts 1 and 3 labeled several hundred particles with a size less than 10 m, while Expert 2 labeled fewer than 100 within this range. The number of particles segmented also tends to correlate with the time spent by each domain expert, which varies from 145 min (Expert 2) to 430 minutes (Expert 3). In contrast, the Mask R-CNN model is applied without manual intervention and requires only 3 minutes to analyze the entire dataset on a desktop CPU. Although it segments fewer particles than two of the experts, specifically missing many of those with a small area, the overall distribution and average particle size matches qualitatively well with the manually crafted labels that take $10\text{--}100 \times$ longer to complete. We also suspect that our model could be refined by collecting additional training data with more labels at small particle size.

Discussion

In this study, we develop segmentation models that can accurately identify distinct particles in powder samples imaged by desktop SEMs. These models can be applied at disparate length scales, to images acquired at low or high magnification, and are robust against measurement artifacts such as image blur. They can also assess their own confidence in the segmentation of each individual particle, which is made possible by modifying the recently developed Mask R-CNN architecture. The ability to gauge a prediction’s accuracy is crucial for the practical use of these models in real-world applications, particularly in materials science and other fields that rely on precise conclusions made from the analysis of characterization data.

When compared to the more traditional U-Net segmentation models, we find that our Mask R-CNN architecture provides improved performance. This is especially true in cases involving blurry and overlapping particles, both of which are highly prevalent in SEM images. The improved performance of the Mask R-CNN models can be traced to their architecture, which allows for the modeling of overlapping instances as isolated entities. This differs from the U-Net models whose architecture was originally designed for semantic segmentation, whereby pixels are assigned to pre-defined categories. In addition, our Mask R-CNN models provide comparable performance to domain experts while requiring substantially less

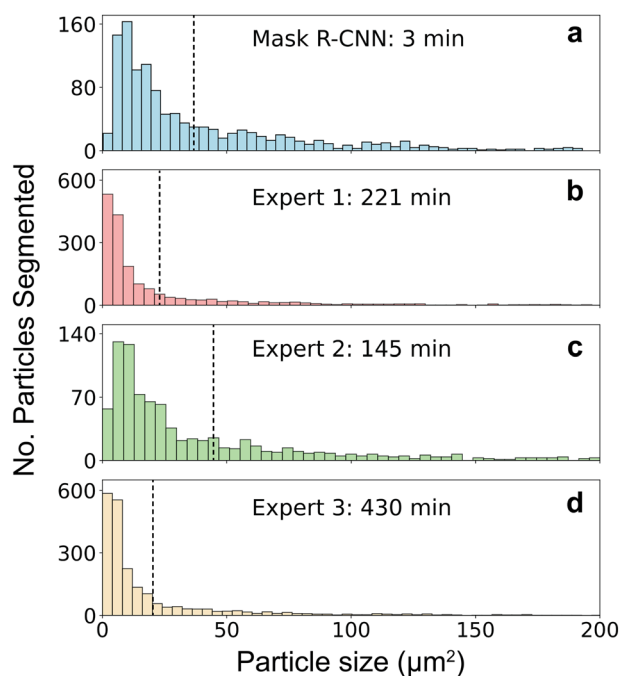


Fig. 6 | Comparison of particle size distribution in LiCoO₂ sample segmented by Mask R-CNN and domain experts. **a** Shows the Mask R-CNN model trained at low magnification, and **b–d** three separate domain experts. The time required by each labeling method is listed near the top of each plot. The average particle size computed from each method is denoted by the vertical dashed lines.

time. The results of our tests on LiCoO₂ showcase the primary advantages of deep learning when applied to segmentation, providing an order-of-magnitude improvement in speed while also mitigating the variability that exists between the labels made by separate domain experts.

Despite their generally positive results, it is worth noting that our model's performance is not without limitations. For instance, the segmentation accuracy becomes more limited when dealing with images captured at particularly high magnification, which often contain blurry particles with loosely defined boundaries, which however is expected since such samples are generally more challenging and ambiguous. Our models also tend to be conservative in their predictions, often segmenting fewer particles than domain experts and overlooking particularly small ones, which can be improved with more data and model regularization techniques. These observations highlight the need for further improvements in this area. Further, to improve upon our method, the threshold value separating certain and uncertain predictions may be optimized in future works to enhance predictive accuracy. Additionally, other out-of-distribution materials besides LiCoO₂ could validate our model's robustness and reveal areas for further improvement.

We believe the results presented in this study have significant implications for future research and applications. They complement recent advancements in laboratory automation, opening up the possibility of integrating SEM/EDS measurements into autonomous workflows for accelerated materials synthesis and characterization^{3,30}. By providing fast and reliable particle segmentation, our models contribute to the development of more efficient and data-driven experiments in materials science and chemistry. Future work might focus on enhancing the performance of segmentation models at high magnifications, addressing the challenges associated with image blur, and refining the models to achieve segmentation results closer to the domain experts. The optimization of these models, and their integration with automated systems for materials characterization, offers the potential for transformative advancements in materials science, accelerating the pace of discovery and innovation in the field.

Methods

Mask R-CNN architecture

Here we explain in detail the Mask R-CNN architecture employed in this work (Fig. 2). Traditional CNNs struggle to handle the diversity of particle sizes prevalent in SEM due to their single-scale feature representations. To address this challenge, Mask R-CNN employs a Feature Pyramid Network (FPN), which enhances feature extraction by providing a multi-scale representation of the input image, enabling effective detection of objects with different sizes. The FPN used in our work leverages multiple intermediate layers of the ResNet-50 backbone to build a feature hierarchy. Starting from the bottom of the features, the FPN uses a series of lateral connections to create feature pyramids. The lateral connections help propagate information from higher-resolution layers to lower-resolution layers, allowing each level of the pyramid to access features from all levels below it. These connections ensure that fine-grained information is preserved at all scales. After that, the FPN samples-up the features from higher levels of the pyramid to match the spatial resolution of features at lower levels. This allows for a smooth fusion of information from different scales. The upsampled features are combined with the lower-level features to get a set of multi-scale feature maps, where each map represents features at a different spatial resolution. These feature maps serve for multi-scale object detection, with high-resolution maps identifying small particles and lower-resolution maps detecting larger ones.

The multi-scale features are used for ROI proposal and alignment. First, the Region-Proposal-Network (RPN) proposes potential regions in the SEM image that may contain objects of interest. In our case, the RPN is a CNN that slides a small window of 3×3 pixels over the feature map and, at each position, predicts multiple rectangular regions (proposals) and their associated objectness scores. The objectness score indicates the likelihood that some region contains an object. After generating region proposals, the RPN uses non-maximum suppression (NMS) to filter out redundant or overlapping proposals and retains the top-ranked ones for further processing. After the ROIs are found, ROI align is used to align and crop the potential particles from the original SEM image at the locations of the region proposals and resize them to a constant size, which is crucial for further processing. In our work, ROI align resizes each region proposal to a fixed spatial size of 7×7 pixels.

The aligned proposals are then processed by multiple modules (heads) responsible for specific tasks. The original Mask R-CNN architecture consists of three heads. The bounding box regression head refines the localizations of the RPN to more accurately fit the objects within the regions. In our case, the bounding box regression head is a two-layer fully connected neural network. The region classification head predicts a class label for each proposed region. Since we do not need a distinction between different classes, this head is discarded in our architecture. The last head of the standard Mask R-CNN architecture is the mask head that generates pixel-wise segmentations for each detected object that outputs a binary mask for each detected object, where each pixel is either classified as belonging to the object or not. In our work, the mask head is a four-layer CNN network. Finally, our estimation of uncertainty is achieved through an additional neural network output head that learns to predict the confidence for each detected particle. In contrast to the objectness, which prioritizes regions for further analysis based on their likelihood of containing actual objects rather than background or noise and acts as a selector and filter for object detection and segmentation, our proposed uncertainty head is an independent component whose output does not serve as a basis for other parts of the network and specifically focuses on discerning between certain and uncertain instances that have been classified as objects of interest by the objectness score, providing a nuanced measure of confidence. We realize this through a single-layer fully connected neural network. To create the necessary labels for this detection, the ground-truth labels of the SEM scan are partitioned into two distinct categories: uncertain particles ψ and certain particles ω . To further refine the estimation of confidence, we introduce noise to each of the true labels to signify that the ground truth may not always be correct. This noise is generated by random sampling from a

Gaussian distribution. Specifically, the mean of this distribution is set at 1 for ω particles and 0 for ψ particles. The standard deviation of this Gaussian distribution is represented as σ . By incorporating this noise, the model conveys that uncertain particles should have label values close to 0, indicating low confidence, while certain particles should be close to 1, signifying a high level of confidence in their detection. The output is a confidence value γ , which is constrained to a range between 0 and 1. Any values of γ that fall below 0 are automatically adjusted to 0, and values exceeding 1 are capped at 1. In essence, this restriction ensures that the confidence value γ consistently falls within the interval $[0, 1]$.

Label uncertainty loss

The loss function calculation of the Mask R-CNN architecture is divided into two components targeting specific aspects of model performance. The first component \mathcal{L}_1 consists of the objectness classification \mathcal{L}_o and the bounding box regression \mathcal{L}_{b_1} . The goal of \mathcal{L}_o is to distinguish between foreground (particles) and background regions in a predicted bounding box and is implemented as a softmax cross-entropy loss. This component does not encompass any uncertainty estimation; it focuses solely on learning to differentiate regions where particles are from all other entities. The objective of \mathcal{L}_{b_1} is to train the model to accurately predict bounding box offsets for the proposed regions. This loss is computed as the Huber loss. Overall, \mathcal{L}_1 is computed as a weighted sum of \mathcal{L}_o and \mathcal{L}_{b_1} : $\mathcal{L}_1 = \mathcal{L}_o + \lambda_{b_1} \mathcal{L}_{b_1}$, where λ_{b_1} is a parameter that specifies the importance of the bounding box regression within the total loss calculation. In our case, λ_{b_1} is always 1. The computation of \mathcal{L}_1 follows the original formulation³¹.

The second component \mathcal{L}_2 consists of three parts: the segmentation loss \mathcal{L}_s , the refined bounding box regression loss \mathcal{L}_{b_2} , and the confidence loss \mathcal{L}_c . The segmentation loss \mathcal{L}_s aims to generate object masks for each ROI. Unlike traditional segmentation architectures such as U-Net, there is no competition among classes when generating masks. Each ROI is treated in isolation, and for each detected object the Binary Cross-Entropy Loss is calculated. Expanding upon \mathcal{L}_{b_1} , the \mathcal{L}_{b_2} further refines the bounding box coordinates for the ROIs to align them more accurately with the ground truth bounding boxes employing the Huber Loss.

Our refinement of the Mask R-CNN architecture \mathcal{L}_c models the uncertainty for each particle as a regression. Our goal is to model a bimodal distribution, with distinct peaks representing high and low uncertainty. However, common loss functions like Mean Absolute Error (MAE), Huber loss, and Euclidean loss are designed for unimodal distributions, not equipped to handle this bimodal uncertainty distribution's complexity. Inspired by approaches that discretize a multimodal distribution into k bins to model the regression as a classification challenge³², we introduce a novel approach that maps the confidence scores $\gamma \in [0, 1]$ to the desired k th bin using the function $\lfloor k\gamma \rfloor$. During training a classification task, we have the batch size $N \in \mathbb{N}$ and $C \in \mathbb{N}$ classes. In our case, the C classes represent the k bins. To formulate the loss function for confidence estimation, we employ a cross-entropy loss extended with a distance regularization term to penalize smaller distances from the ground truth less. This loss function is defined as:

$$\mathcal{L}_c(\mathbf{x}, \mathbf{y}, \xi) = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C w_c \log \frac{\exp(x_{n,c})}{\sum_{i=1}^C \exp(x_{n,i})} \exp\left(\frac{-|y_n - c|}{\xi}\right), \quad (1)$$

where \mathbf{x} represents the model's prediction, with a distribution for each class, \mathbf{y} is the target containing class indices, \mathbf{w} is a weight vector specifying the importance for each class, $|y_n - c|$ represents the absolute difference between the class index y_n and the class c , and ξ is a parameter controlling how much similarity between \mathbf{x} and \mathbf{y} is enforced. Like this, we map confidence scores to bins, allowing the model to handle multimodal distributions effectively and penalize uncertainties with different degrees. The calculation of the ground truth bins is visualized in Fig. 7.

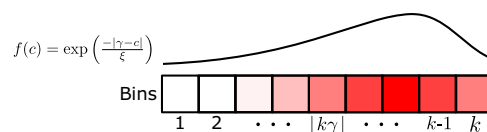


Fig. 7 | Visual depiction of how the ground-truth bins for \mathcal{L}_c are calculated. γ is the certainty value, c the observed class, and ξ is the similarity-enforcing factor. The function is discretized into k bins.

\mathcal{L}_2 is calculated as a sum of the individual parts $\mathcal{L}_2 = \mathcal{L}_s + \mathcal{L}_{b_2} + \mathcal{L}_c$, and the overall loss \mathcal{L} as the sum of both components: $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$. The parameters for the loss calculation are optimized in a hyperparameter search (Supplementary Table 3).

Combination of masks

The Mask-RCNN output consists of m independent masks, one mask for each detected particle. Multiple post-processing steps are utilized to create a unified segmentation mask. To output certain particles ω , the m masks are filtered to only contain masks where the maximum value of the k bins is greater or equal to $k/2$. Conversely, to obtain the uncertain particles ψ , only those in m where the maximum confidence is smaller than $k/2$ are observed. Finally, if no distinction is made, m is not filtered. In the next step, the retained masks are sorted based on their predicted objectness scores \hat{o} in descending order. Subsequently, masks with \hat{o} smaller than specified thresholds θ_ω for ω masks and θ_ψ for ψ masks are removed. We then iterate through the remaining masks, comparing each mask with all subsequent masks. If the Intersection-Over-Union (IoU) score between two compared masks exceeds a predetermined threshold θ_\cap , we discard the second one to remove masks that have been detected twice. In cases where the final prediction should encompass both ω and ψ masks, an additional post-processing step is carried out. Specifically, we compute the IOU score for each pair of masks from the ω and ψ categories and if the IOU score for a given particle pair surpasses a threshold θ_\cap , the ψ mask is removed to ensure that the final prediction is not compromised by overlapping information in which a particle is both classified as certain and as uncertain. Finally, the segmentation masks are binarized using an activation threshold θ_Σ . The parameters for the post-processing are optimized in a hyperparameter search (Supplementary Table 4).

Training details

We enhance the original U-Net to contain a ResNet-50 backbone to employ the same encoder for both models thus increasing comparability. Our U-Net implementation consists of 38.5 million trainable parameters. The Mask R-CNN contains 44.0 million trainable parameters, from which 6,150 originate from our novel uncertainty head. We use the Adam optimizer with a learning rate of 0.005 for the U-Net model, while the Mask R-CNN utilizes an AdamW optimizer with a learning rate of 0.0001. All samples are normalized to be in the range $[0, 1]$. Early stopping and learning rate scheduling are employed. Both U-Net and Mask R-CNN are implemented in PyTorch Lightning. We conduct training on the super-computer system Hochleistungsrechner Karlsruhe (HoreKa) at KIT. One computational node of HoreKa is equipped with an Intel Xeon Platinum 8368 CPU (2 sockets, 76 cores per socket) and four NVIDIA A100 Tensor Core GPUs. The analysis of the LiCoO₂ powder for comparing the runtime against human experts was conducted on an AMD Ryzen 9 5950x CPU. To avoid initialization effects and ensure reliable metrics, we repeat all experiments four times. We do not resize the SEM scans for training to ensure that all details in the images remain identifiable. Due to the large image size, the mini batch size is 1 for both the U-Net and Mask R-CNN models. For the uncertainty estimation (Subsection 2.4), the ground truth bounding boxes are provided to the Mask R-CNN to solely evaluate the confidence predictions. Further, since our model outputs the confidence as a decimal, we regard a prediction as certain if the confidence is $>50\%$ and uncertain if it is $\leq 50\%$.

Data availability

All code used here can be found in the GitHub repository <https://github.com/lrettenberger/Uncertainty-Aware-Particle-Segmentation-for-SEM>. The data used to train the models are available at: https://figshare.com/projects/Uncertainty-aware_particle_segmentation_for_electron_microscopy_at_varied_length_scales/201381 and <https://github.com/lrettenberger/Uncertainty-Aware-Particle-Segmentation-for-SEM-Data>.

Received: 5 December 2023; Accepted: 24 May 2024;

Published online: 12 June 2024

References

- Patil, S. et al. Alternate synthesis method for high-performance manganese rich cation disordered rocksalt cathodes. *Adv. Energy Mater* **13**, 2203207 (2023).
- Li, W. et al. Peering into batteries: electrochemical insight through in situ and operando methods over multiple length scales. *Joule* **5**, 77–88 (2021).
- Szymanski, N. J. et al. Toward autonomous design and synthesis of novel inorganic materials. *Mater. Horiz.* **8**, 2169–2198 (2021).
- Hill, J. et al. Materials science with large-scale data and informatics: Unlocking new opportunities. *MRS Bull.* **41**, 399–409 (2016).
- Park, C., Huang, J. Z., Ji, J. X. & Ding, Y. Segmentation, inference and classification of partially overlapping nanoparticles. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 669–681 (2012).
- 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* **9**, 12488–12495 (2015).
- Lee, B. et al. Statistical characterization of the morphologies of nanoparticles through machine learning based electron microscopy image analysis. *ACS Nano* **14**, 17125–17133 (2020).
- Bell, C. G. et al. Trainable segmentation for transmission electron microscope images of inorganic nanoparticles. *J. Microsc.* **288**, 169–184 (2022).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing Computer-Assisted Intervention* (eds. Navab, N. et al.) 234–241 (Springer, 2015).
- Yang, S.-H. et al. Deep learning-assisted quantification of atomic dopants and defects in 2d materials. *Adv. Sci.* **8**, 2101099 (2021).
- Horwath, J. P., Zakharov, D. N., Mégrét, R. & Stach, E. A. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *npj Comput. Mater* **6**, 108 (2020).
- Ghosh, A. et al. Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy. *npj Comput. Mater* **7**, 100 (2021).
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. Unet+: A nested u-net architecture for medical image segmentation. In *Proc. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds. Stoyanov, D. et al.) 3–11 (Springer, Cham, 2018).
- Stuckner, J., Harder, B. & Smith, T. M. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *npj Comput. Mater* **8**, 200 (2022).
- Choudhary, K. et al. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater* **8**, 59 (2022).
- Rettenberger, L., Rieken Münke, F., Bruch, R. & Reischl, M. Mask r-cnn outperforms u-net in instance segmentation for overlapping cells. *Curr. Directions Biomed. Eng.* **9**, 335–338 (2023).
- Ragone, M., Shahabzian-Yassar, R., Mashayek, F. & Yurkiv, V. Deep learning modeling in microscopy imaging: a review of materials science applications. *Prog. Mater. Sci.* **138**, 101165 (2023).
- Jacobs, R. Deep learning object detection in materials science: Current state and future directions. *Comput. Mater. Sci.* **211**, 111527 (2022).
- Agarwal, S. et al. Comparing u-net and mask r-cnn algorithms for deep learning-based segmentation of electron microscopy images containing cavities for nuclear reactor applications. In *Proc. International Conference On Electrical, Computer, Communications and Mechatronics Engineering* 1–4 (IEEE, 2023).
- Cohn, R. et al. Instance segmentation for direct measurements of satellites in metal powders and automated microstructural characterization from image data. *JOM* **73**, 2159–2172 (2021).
- Price, S. E., Gleason, M. A., Sousa, B. C., Cote, D. L. & Neamtu, R. Automated and refined application of convolutional neural network modeling to metallic powder particle satellite detection. *Integr. Mater. Manuf. Innov.* **10**, 661–676 (2021).
- Boyle, M. J., Goldman, Y. E. & Composto, R. J. Enhancing nanoparticle detection in interferometric scattering (ISCAT) microscopy using a mask r-cnn. *J. Phys. Chem.* **127**, 3737–3745 (2023).
- Lin, B. et al. A deep learned nanowire segmentation model using synthetic data augmentation. *npj Comput. Mater* **8**, 88 (2022).
- Jacobs, R. et al. Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images. *Sci. Rep.* **13**, 5178 (2023).
- Chen, Q., Zheng, C., Cui, Y., Lin, Y.-R. & Zinkle, S. J. A deep learning model for automatic analysis of cavities in irradiated materials. *Comput. Mater. Sci.* **221**, 112073 (2023).
- Ghomrasni, N. B., Chivas-Joly, C., Devoille, L., Hochepped, J.-F. & Feltin, N. Challenges in sample preparation for measuring nanoparticles size by scanning electron microscopy from suspensions, powder form and complex media. *Powder Technol.* **359**, 226–237 (2020).
- Botifoll, M., Pinto-Huguet, I. & Arbiol, J. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. *Nanoscale Horiz.* **7**, 1427–1477 (2022).
- Kumar, N. et al. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* **36**, 1550–1560 (2017).
- Scherr, T., Löffler, K., Böhländ, M. & Mikut, R. Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy. *PLoS ONE* **15**, 1–22 (2020).
- Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 1476–4687 (2023).
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. In *Proc. IEEE International Conference on Computer Vision* 2961–2969 (IEEE, 2017).
- Massa, F., Marlet, R. & Aubry, M. Crafting a multi-task cnn for viewpoint estimation. In: *Proc. British Machine Vision Conference*, 91.1–91.12 (British Machine Vision Association (BMVA), 2016).

Acknowledgements

The experiments performed in this work were financed by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05-CH11231 (D2S2 programme, KCD2S2). N.J.S. and S.W. were supported in part by the Jane Lewis Fellowship at UC Berkeley. This work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

Author contributions

All authors designed the research. Luca Rettenberger and Nathan J. Szymanski developed the methods. Luca Rettenberger implemented and conducted the experiments. Markus Reischl & Gerbrand Ceder conceived

and supervised the project. Nathan J. Szymanski, Yan Zeng, and Shilong Wang recorded and annotated the datasets. All participated in data analysis. Luca Rettenberger wrote the paper, supported by Nathan J. Szymanski and Markus Reischl.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-024-01302-w>.

Correspondence and requests for materials should be addressed to Markus Reischl.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024