



OPEN

# A deep learning approach for semantic segmentation of unbalanced data in electron tomography of catalytic materials

Arda Genc<sup>1,3</sup>, Libor Kovarik<sup>2</sup>✉ & Hamish L. Fraser<sup>1</sup>

In computed TEM tomography, image segmentation represents one of the most basic tasks with implications not only for 3D volume visualization, but more importantly for quantitative 3D analysis. In case of large and complex 3D data sets, segmentation can be an extremely difficult and laborious task, and thus has been one of the biggest hurdles for comprehensive 3D analysis. Heterogeneous catalysts have complex surface and bulk structures, and often sparse distribution of catalytic particles with relatively poor intrinsic contrast, which possess a unique challenge for image segmentation, including the current state-of-the-art deep learning methods. To tackle this problem, we apply a deep learning-based approach for the multi-class semantic segmentation of a  $\gamma$ -Alumina/Pt catalytic material in a class imbalance situation. Specifically, we used the weighted focal loss as a loss function and attached it to the U-Net's fully convolutional network architecture. We assessed the accuracy of our results using Dice similarity coefficient (DSC), recall, precision, and Hausdorff distance (HD) metrics on the overlap between the ground-truth and predicted segmentations. Our adopted U-Net model with the weighted focal loss function achieved an average DSC score of  $0.96 \pm 0.003$  in the  $\gamma$ -Alumina support material and  $0.84 \pm 0.03$  in the Pt NPs segmentation tasks. We report an average boundary-overlap error of less than 2 nm at the 90th percentile of HD for  $\gamma$ -Alumina and Pt NPs segmentations. The complex surface morphology of  $\gamma$ -Alumina and its relation to the Pt NPs were visualized in 3D by the deep learning-assisted automatic segmentation of a large data set of high-angle annular dark-field (HAADF) scanning transmission electron microscopy (STEM) tomography reconstructions.

TEM tomography in materials science has become the de facto technique that enables valuable high spatial resolution information on the structure of the materials in 3D<sup>1–5</sup>. Despite the progress in developing novel methods for acquiring and aligning the tomography tilt series and a broad spectrum of reconstruction algorithms<sup>6–9</sup>, semantic segmentation of the large 3D data sets remains a significant bottleneck in 3D analysis. Manual segmentation is a very time-consuming task, relying heavily on the handcrafting skills and expertise of a human operator. A reliable, reproducible, and fully automated segmentation method is in high demand for scaling the 3D data analysis and collecting statistically meaningful information where a single reconstructed volume segmentation involves hundreds of images or more.

Recent advances in deep learning methods have revolutionized the field of computer vision<sup>10–13</sup>, and the evolution of these methods has enabled the automatic semantic segmentation of large data sets; otherwise, manual analysis is unfeasible<sup>14–16</sup>. Deep learning pixel-wise classifiers have been successfully applied to many semantic segmentation tasks where complex structures are not easily mapped by simple intensity differences, and boundaries between the image features are not apparent due to the variations in contrast gradients<sup>17</sup>.

In deep learning, fully convolutional neural networks (FCNs) hierarchically recognize complex features directly from the training data without the additional feature engineering. More recently, FCNs, inspired by large and deep networks, are efficiently trained end-to-end by supervised learning and pixel-to-pixel probabilities computed successfully, thanks to the many advancements in parallel computing<sup>18</sup>. It has been shown that

<sup>1</sup>Center for the Accelerated Maturation of Materials, Department of Materials Science and Engineering, The Ohio State University, Columbus, OH, USA. <sup>2</sup>Institute for Integrated Catalysis, Pacific Northwest National Laboratory, Richland, WA, USA. <sup>3</sup>Present address: Materials Department, University of California Santa Barbara, Santa Barbara, CA, USA. ✉email: libor.kovarik@pnnl.gov

segmentation results using FCNs can indeed reach a human-level performance, and even on some occasions, exceed that without the post-tuning of the results<sup>19–21</sup>. Today, it is possible to generalize these deep networks with a limited amount of ground-truth data by implementing data augmentation and regularization techniques while mitigating the problem of variance<sup>22</sup>.

In this paper, we explore the capability of FCNs in the semantic segmentation of a  $\gamma$ -Alumina ( $\text{Al}_2\text{O}_3$ )/Pt catalytic material. Historically,  $\gamma$ -Alumina has been one of the most used catalytic support materials for noble metals and oxide catalysts employed for reduction, oxidation, and reforming reactions in automotive exhaust control and petroleum refining processes<sup>23</sup>.  $\gamma$ -Alumina possesses a complex crystalline structure; despite its broad application space, the origin of the catalytic behavior remains actively studied<sup>24,25</sup>. There is a considerable debate on the role of surfaces of  $\gamma$ -Alumina responsible for both catalytic properties and anchoring of the noble metallic NPs. In addition to the structural complexity and small crystallite sizes, the support material  $\gamma$ -Alumina consists of a dense network of matrix pores, and the degree to which these pores are connected to outside surfaces is of great interest<sup>24–27</sup>.

In catalytic materials, the sparse distribution of the noble metallic NPs, over the background and oxide support material introduces an unbalanced representation of the data in TEM images. The extent of the class imbalance problem between the foreground and background of the images has been extensively studied in deep learning-based semantic segmentation approaches<sup>28–30</sup>. It has been shown that the choice of loss function significantly impacts the performance of a semantic segmentation model<sup>31</sup>. Many recent state-of-the-art applications of FCNs focus on the implementation of weighting strategies coupled with distribution-based or differentiable region-based loss functions for the optimization of the models<sup>31–33</sup>. Even though weighting strategies at the loss function level control the class imbalance, the problem of loss becoming overwhelmed by the number of easy examples during inference remains a challenge in complex multi-class situations<sup>34</sup>. Moreover, there are limited applications of these strategies in the semantic segmentation of materials science samples, particularly segmentation of the 3D electron tomography reconstructions<sup>15,21,35,36</sup>. This study presents the first application of deep learning-based multi-class semantic segmentation of large and unbalanced data of 3D tomograms as obtained from HAADF STEM.

## Results and discussion

We present a U-Net-based FCN architecture and weighted focal loss as a loss function to overcome the class imbalance problem<sup>34,37</sup>. The weighted focal loss is a distribution-based loss function, and weighting of the unbalanced data occurs at the loss function level in contrast to data preprocessing strategies. In addition to the weighting, focal loss applies a modulation term to the standard cross-entropy loss and dynamically scales the confidence of the correctly classified examples<sup>34</sup>. In our experiments, the U-Net architecture equipped with the weighted focal loss facilitated a comprehensive 3D representation of the catalytic material and provided a clear insight regarding the long-standing debate on the characteristics of  $\gamma$ -Alumina surfaces and their relation to the catalytic NPs. We discuss the accuracy of our segmentation results by assessing commonly used semantic segmentation metrics on the overall overlap and boundary match between the ground-truth and predicted segmentations.

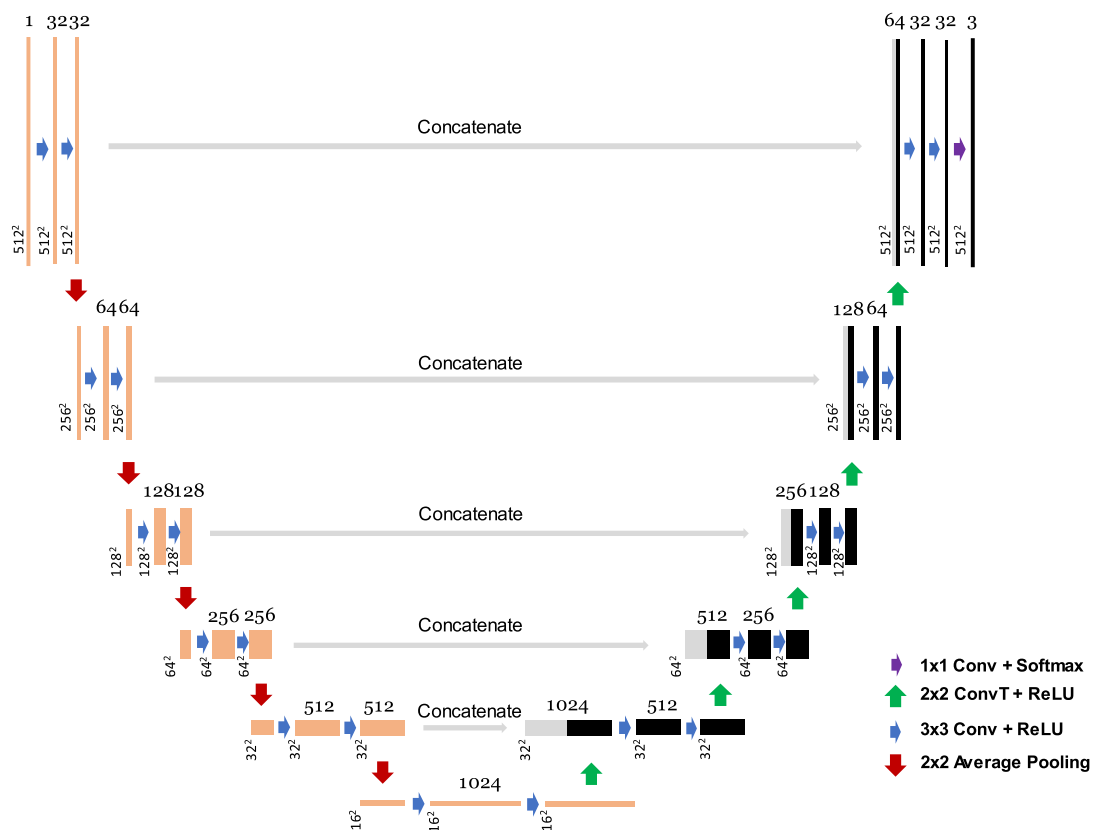
To further test our model's robustness and validity, the best-performing model was deployed on the automatic semantic segmentation of a large data set of reconstructed images. We believe strongly that deep learning-based semantic segmentation methods have immense potential in 3D data analysis and will usher in a new era in materials design and discovery.

**Segmentation architecture and inference.** For FCNs experiments, we exploit an adopted version of the U-Net architecture, and a schematic of the architecture is shown in Fig. 1. Our network consists of two learning paths, a down-sampling (contraction) path and an up-sampling (expansion) path. There are six convolutional steps in the down-sampling path and five in the up-sampling path. In the down-sampling path, each step has two convolutional layers with a filter size of  $3 \times 3$ . The size of the feature maps is halved by the pooling layers following each step. In the up-sampling path, each step starts with a convolutional transpose layer with a filter size of  $2 \times 2$  and a stride of 2, followed by two convolutional layers with a filter size of  $3 \times 3$ . The size of the feature maps is doubled, and the number of feature maps is halved at the end of each convolutional step in the up-sampling path.

In our version of the U-Net, we used the 'same' padding in the convolution layers followed by an average pooling layer for down-sampling. Using the 'same' padding resulted in the output of the convolution layers being the same size as the input layers. The pooling operation plays a vital role in the flow of information through the convolutional layers and defines the model's sensitivity to details. In our architecture, average pooling is preferred over max pooling for down-sampling to reduce the spatial information loss at the feature boundaries and prevent excessive pixel saturation.

Concatenation paths (skip connections) give the network a well-known 'U' shape pattern and link the high spatial information from down-sampling convolutional layers to the up-sampling convolutional layers. The network hierarchically learns the contextual information and fine details in the predicted images. We did not see a significant performance improvement in our model with the addition of dropout layers; instead, we employed a comprehensive data augmentation strategy for regularization. We used a rectified linear unit (ReLU) activation function for hidden layers and softmax for the output convolutional layer with final feature maps of 3.

As described previously, we investigate the complex bulk and surface structure of the  $\gamma$ -Alumina/Pt catalysts. Our segmentation model aims to distinguish the two-phase microstructure of the  $\gamma$ -Alumina and Pt NPs, as well as the pores. A correct classification of the surfaces, defined by the  $\gamma$ -Alumina—background and  $\gamma$ -Alumina—Pt—background boundaries, is mainly of high interest. In the tomography reconstructions,  $\gamma$ -Alumina and the background constitute the most significant fraction of the reconstructions, compared with the sparsely



**Figure 1.** A schematic of the U-Net architecture. The number of feature maps is indicated on the top of each box, and the dimensions of the feature maps are on the bottom left corner. Orange boxes show the contraction path, and black boxes show the expansion path. The gray arrows and boxes indicate the concatenation path at each convolutional step.

distributed nanoscale Pt particles. When there is an imbalance in the data representation, the learning algorithm can be biased towards the dominant class, represented densely by the higher number of pixels. In order to address this problem, we used weighted focal loss instead of the standard cross-entropy to minimize the loss<sup>34,38</sup>. Weighted focal loss is a differentiable modification of the cross-entropy loss term and addresses the class imbalance problem in two ways. Firstly, it shifts the focus from easy-to-classify pixels towards hard-to-classify pixels by extending the range in which each pixel receives loss. This is achieved by scaling the cross-entropy loss by a focusing parameter and preventing the loss from being overwhelmed by the easy pixels. Secondly, the role of the weighted focal loss is to provide a balancing action on the class imbalance problem by adjusting the contribution of each class to the loss function by a weighting factor. Categorical focal loss ( $L_{FL}$ ) is defined as the following in a multi-class problem:

$$L_{FL}(y, p) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \alpha_c y_{i,c} (1 - p_{i,c})^\gamma \log p_{i,c} \tag{1}$$

where  $y_{i,c}$  and  $p_{i,c}$  are the ground-truth and prediction probabilities of class  $c$  at pixel location  $i$ . Parameters  $C$  and  $N$  are the number of classes and pixels, respectively.  $\alpha_c$  is the weighting factor for class  $c$  and  $\gamma$  is the focusing parameter. Both focusing parameter and weighting factor are tunable hyperparameters. In our experiments,  $\alpha_c$  values were approximated on the density of representation of each class at the range from 0 to 1, and  $\gamma$  was set to 1.

**Evaluation metrics.** Evaluation metrics play an essential role in proving the network’s performance and thus establishing the model for automatic semantic segmentation. In this work, predictions were assessed using four commonly used semantic segmentation metrics: Dice similarity coefficient (DSC), recall, precision, and Hausdorff distance (HD)<sup>39</sup>. DSC, recall, and precision scores are similarly extracted from the confusion matrix and defined as:

$$DSC = \frac{2TP}{2TP+FP+FN} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

where true positives (TP), false positives (FP), and false negatives (FN) represent per pixel classifications of the confusion matrix. In this study, a  $3 \times 3$  confusion matrix was calculated from each ground-truth and predicted segmentations at pixel resolution of 0.12 nm/pixel using the validation data set. Average DSC, recall, and precision scores were reported using a mean  $\pm$  95% confidence interval.

We also evaluated our semantic segmentation results by measuring the dissimilarities specifically at the segmentation boundaries. Hausdorff distance is a boundary distance-based metric and measures the largest segmentation error in the overlap between the ground-truth and predicted segmentations<sup>40</sup>. Given two sets of points A and B, Hausdorff distance is defined as:

$$HD(A, B) = \max(hd(A, B), hd(B, A)) \quad (5)$$

where  $hd(A, B)$  and  $hd(B, A)$  are directed Hausdorff distances:

$$hd(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (6)$$

$$hd(B, A) = \max_{b \in B} \min_{a \in A} \|a - b\| \quad (7)$$

Functions  $hd(A, B)$  and  $hd(B, A)$  measure the distances between two points in A and B, which are farthest from any nearest neighbors, and  $HD(A, B)$  gives the largest of these distances.  $\|a - b\|$  is the Euclidean norm between the points in A and B. A well-documented behavior of the Hausdorff distance is its sensitivity to outliers and noise<sup>41</sup>; thus, we report robust HD (RHD) values considering the percentile of the largest segmentation errors and as well as the maximum HD<sup>42,43</sup>. We aim to down-weight the impact of outliers and noise on the HD metric by measuring the RHD values.

**Classification of the 3D reconstructions.** Segmentation of the HAADF STEM tomography reconstructions is a challenging task due to information loss from the insufficient number of projections (i.e., missing wedge artifacts) and variations in the contrast and size of the features in tilt images. A series of representative orthogonal slices (orthoslices) taken from the reconstructed 3D volume of an isolated  $\gamma$ -Alumina/Pt particle is shown in Fig. 2a–c. In the orthoslices, the 3D microstructure of the particle is sectioned perpendicular and parallel to the broad surfaces of  $\gamma$ -Alumina. At first glance, we notice the significant contrast difference between the oxide  $\gamma$ -Alumina and metallic Pt NPs, where Pt NPs appear much brighter than both  $\gamma$ -Alumina substrate and background in the reconstructions. Pt NPs mainly exhibit round shapes, and their size distribution is in the 1–4 nm range, while  $\gamma$ -Alumina particles have a thin plate-like structure and a roughly rhombus shape. Some of the very small Pt NPs show an elongated shape because of the well-documented missing wedge artifact in TEM tilt tomography<sup>7</sup>, as seen in Fig. 2c.

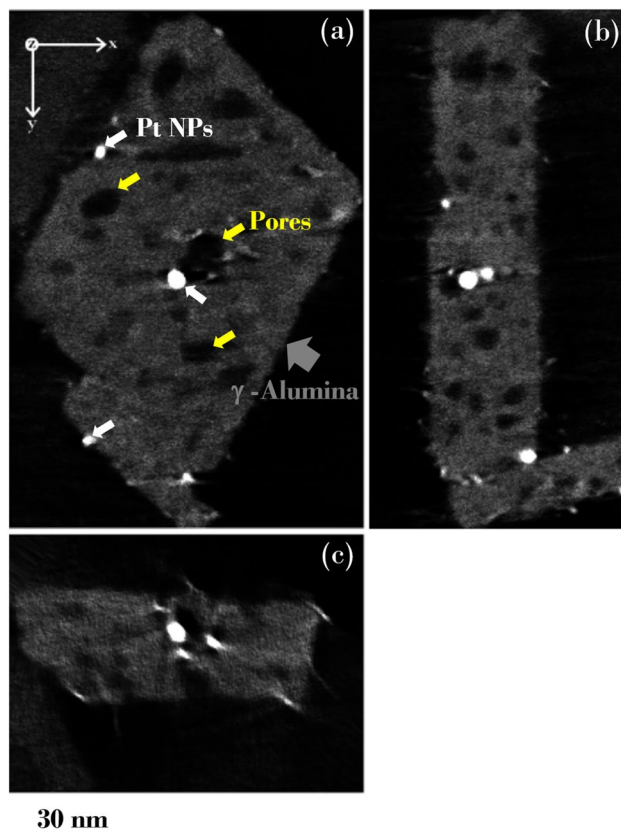
In  $\gamma$ -Alumina, the crystallographic shape of the particle is defined by the {110} and {111} type surfaces, and the main 'broad' surface is {110} orientation, and side surfaces are {111}. Furthermore, synthesized  $\gamma$ -Alumina particle shows small ledges and facets on the surfaces and a network of high-density pores inside the matrix.

To evaluate the performance of our model in the segmentation task, we compared the results obtained from the validation data. Figure 3a,b, visualize results of exemplar qualitative segmentations from  $\gamma$ -Alumina and Pt NPs, respectively. The images represent two  $512 \times 512$  pixels patches from the validation data. For each patch of the validation data, ground-truth and predicted segmentations of  $\gamma$ -Alumina and Pt NPs are compared separately in the binary images. The differences in the segmentations are shown more explicitly in the false negative and false positive maps highlighting the discrepancies in the overlap of each class.

When considering the false positive maps of the  $\gamma$ -Alumina segmentations, we notice that  $\gamma$ -Alumina boundaries are relatively shifted towards the pores inside the  $\gamma$ -Alumina matrix. In contrast to  $\gamma$ -Alumina/pore boundaries, most misclassifications near the surfaces are discontinuous. False negative and false positive regions are extended only a few pixels wide towards either  $\gamma$ -Alumina or background. Moreover, the boundaries in the predicted segmentations, both on the surfaces and along the  $\gamma$ -Alumina/pore boundaries, are smoother than the ground-truth segmentations.

We further discuss these observations in the context of the evaluation metrics, and the results are shown in Table 1. The overlap performance between the ground-truth and predicted segmentations was evaluated using the validation data set consisting of 66 isolated Pt NPs with an average diameter of 16 pixels, and 70  $\gamma$ -Alumina matrix pores with an average diameter of 37 pixels. As expected, shifting of the  $\gamma$ -Alumina boundaries, mainly towards the pores, is reflected in the evaluation results such that the average precision score is lower than the average recall score for  $\gamma$ -Alumina. The average precision and recall scores are  $0.95 \pm 0.008$  and  $0.97 \pm 0.004$  (mean  $\pm$  95% confidence interval), respectively. In contrast to  $\gamma$ -Alumina, we observed fewer false positives in the segmentations of Pt NPs. Still, there are some missing Pt NPs in the predictions and corresponding misclassified regions seen in the comparison of the Pt NPs segmentations. Compared with  $\gamma$ -Alumina, the average precision score of Pt NPs is higher than the recall score. The average precision and recall scores are  $0.92 \pm 0.03$  and  $0.78 \pm 0.04$ , respectively, for the segmentation of Pt NPs.

One explanation for the extension of the  $\gamma$ -Alumina matrix towards the pores inside may be the contrast modulations along the diffuse boundaries between the  $\gamma$ -Alumina and pores. Uncertainty in the contrast of these boundaries can potentially fuel ambiguity in the manual annotations of the ground-truth segmentations. Still, a precise annotation of these low contrast boundaries can be a challenging task even for a human expert.



**Figure 2.** Orthogonal slices are taken from the 3D HAADF STEM reconstructions of an isolated  $\gamma$ -Alumina/Pt particle. (a) A representative section parallel to the broad surface of  $\gamma$ -Alumina, axial slice parallel to XY-plane, (b,c) are axial slices parallel to YZ and XZ-planes, respectively. The large particle with dimmer contrast is  $\gamma$ -Alumina, marked with a gray arrow. Smaller round particles with brighter contrast are Pt NPs, and a few examples are marked with white arrows. Pores are marked with yellow arrows.

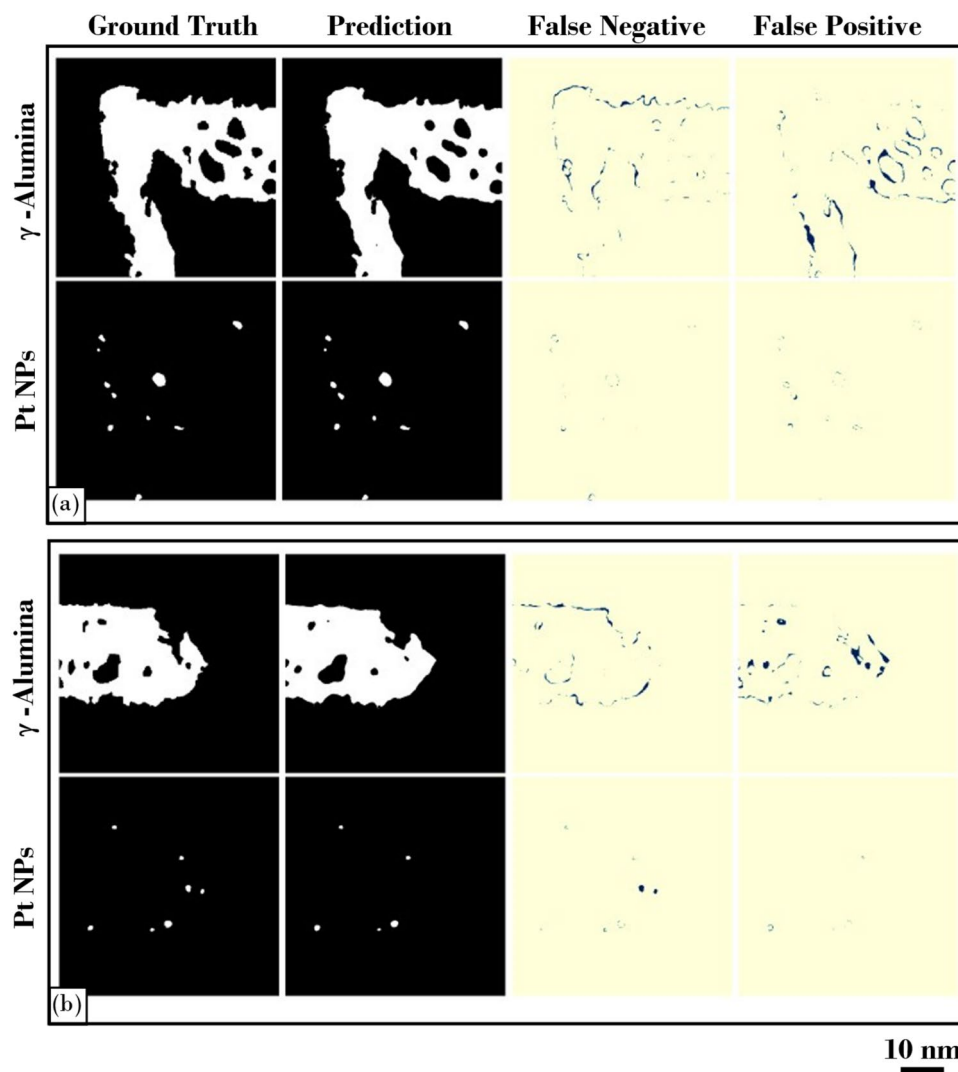
Another approach would be generating more annotated data to elevate the model's performance. However, this would be computationally costly and would require additional manual annotations. It is also worth mentioning that it is inevitable to have false negatives and false positives during inference. Ideally, we would target a trade-off between precision and recall. Nevertheless, a comparison of the ground-truth segmentations with the predicted segmentations shows a strong correlation, especially in the complex surface structure of the  $\gamma$ -Alumina and the appearance of the Pt NPs. The overall similarities in the size and shape of the Pt NPs between the ground-truth and predicted segmentations also suggest that the model has convincingly managed the class imbalance problem without a significant underestimation. The overall segmentation performance measured by the DSC score for each class is  $0.99 \pm 0.002$  for background/pores,  $0.96 \pm 0.003$  for  $\gamma$ -Alumina, and  $0.84 \pm 0.03$  for Pt NPs. Here, we report the background DSC score since the pores inside  $\gamma$ -Alumina are associated with the background class.

To investigate our segmentation results further, we conducted measurements on the degree of boundary match using the HD metric. This is of particular interest for detecting the model's performance correctly learning the boundaries in each class. The HD metric is a powerful tool in measuring the largest segmentation error in the overlap between the two segmentations, while the DSC score is an overlap metric affected by the segmentation performance over the entire image. Figure 4 shows the trend in the average HD of  $\gamma$ -Alumina and Pt NPs segmentations at the various percentiles of robust HD (RHD) values. Measurement of the RHD provides a unique opportunity to understand the contributions of the outliers and noise to the model performance while guiding the degree of boundary match.

As seen in Fig. 4, RHD values fall sharply from a maximum HD of  $5.45 \pm 1.78$  nm (mean  $\pm$  standard error) for  $\gamma$ -Alumina and  $3.70 \pm 1.84$  nm for Pt NPs. At RHD95, HD between the ground-truth and predicted segmentations decreases to  $1.45 \pm 0.68$  nm for  $\gamma$ -Alumina and  $3.48 \pm 1.80$  nm for Pt NPs, and at RHD90, the largest segmentation errors are less than 2 nm ( $0.59 \pm 0.05$  nm for  $\gamma$ -Alumina and  $1.81 \pm 1.05$  nm for Pt NPs). The variations in the RHD values show that the largest segmentation error for the Pt NPs trend higher than for the  $\gamma$ -Alumina, which is consistent with the overall lower evaluation scores observed for the Pt NPs, particularly the lower average recall score. Yet, this analysis suggests that our model learned reasonably well to classify the pixels at the boundaries, especially those of the  $\gamma$ -Alumina in addition to the matrix regions.

Based on the evaluation results, we deployed the best-performing model for the automatic semantic segmentation of a large volume of 3D reconstructions. Figure 5 compares the example reconstructions extracted at



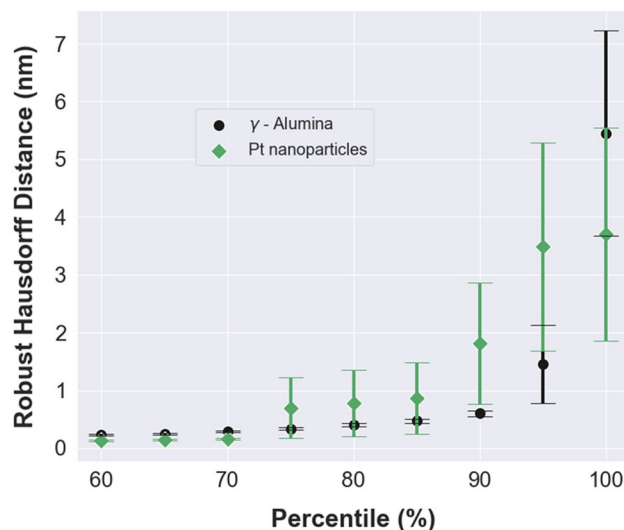


**Figure 3.** A set of ground truth and predicted segmentations of  $\gamma$ -Alumina and Pt NPs from the validation data and corresponding false negative and false positive maps show the discrepancies in the overlap of each class. The images in (a) and (b) represent two  $512 \times 512$  pixels size patches from the validation data set; for illustrative purpose, each set of images demarcated with a black frame. Upper rows are segmentation results from the  $\gamma$ -Alumina particle and lower rows are from the Pt NPs of the same patch.

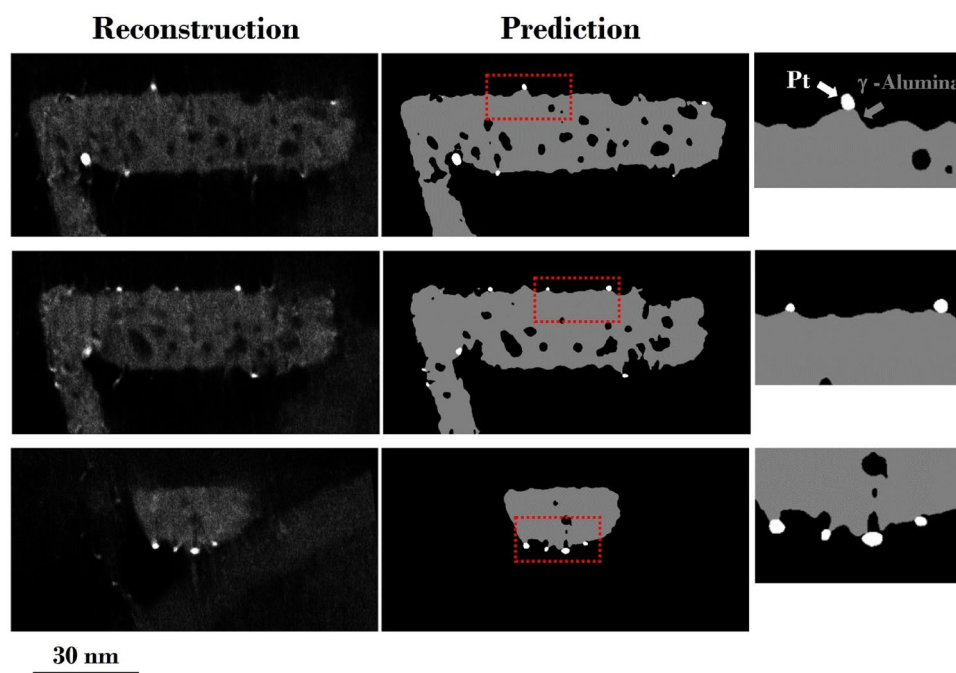
Evaluation metric	$\gamma$ -Alumina	Pt NPs	Bkgnd/pores
Precision	$0.95 \pm 0.008$	$0.92 \pm 0.03$	$0.99 \pm 0.001$
Recall	$0.97 \pm 0.004$	$0.78 \pm 0.04$	$0.99 \pm 0.003$
DSC	$0.96 \pm 0.003$	$0.84 \pm 0.03$	$0.99 \pm 0.002$

**Table 1.** Evaluation results from the validation data set. All the values are in the form of mean  $\pm$  95% confidence interval.

various locations along the 3D volume of the  $\gamma$ -Alumina/Pt particle and corresponding predicted segmentations by the model. The model has not seen these representative reconstruction slices during the training and validation steps. In the predicted segmentation, the pixels classified as  $\gamma$ -Alumina are denoted in gray, Pt in white, and background/pores in black. Overall, there is a good correspondence between the reconstruction slices and predicted segmentations. The location, shape, and size of the Pt NPs correlate with the test images and the texture of the pores inside the  $\gamma$ -Alumina matrix. One striking observation is that the catalytic Pt NPs are associated with the crystallographic modulations on the broad  $\{110\}$  surfaces of the  $\gamma$ -Alumina particle. Most of the Pt NPs are found at the apex of the two  $\{111\}$  type facets, as visualized explicitly in the zoomed-in images.

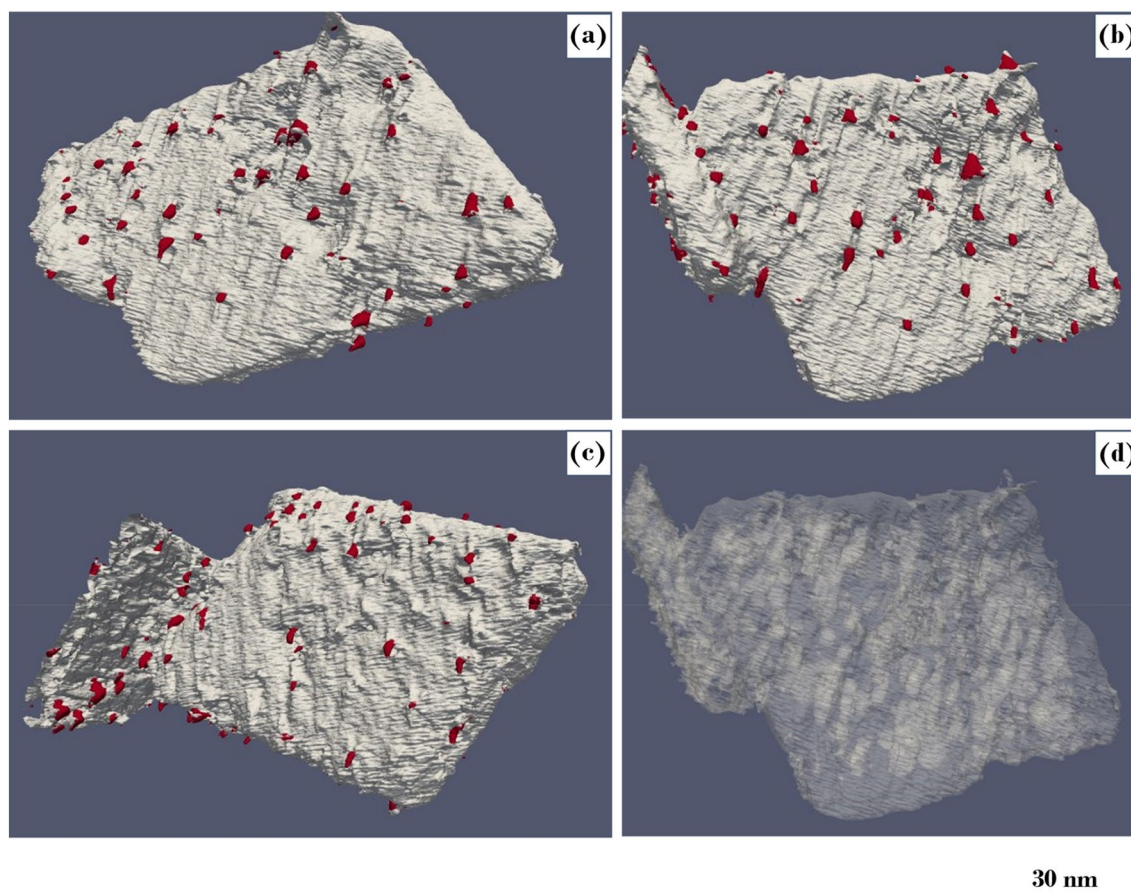


**Figure 4.** A plot of robust Hausdorff distance (RHD) vs. percentile of the largest segmentation error. Data points are in the form of mean  $\pm$  standard error.



**Figure 5.** Comparison of the reconstruction slices (3D reconstructions) and predicted segmentations of  $\gamma$ -Alumina catalytic support material and Pt NPs. Reconstruction slices were extracted at various locations along the 3D volume.

Our model aims to accurately segment the complex surface and bulk microstructures of the  $\gamma$ -Alumina particle and Pt NPs with a limited amount of annotated ground-truth data. The results presented establish that the U-Net model with a weighted focal loss provides a stable model for the multi-class semantic segmentation of a large data set of 3D reconstructions in a severe class imbalance situation. Segmentation results establish a basis for the quantification of critical microstructural parameters, including 1) quantification of external surfaces in terms of their general area and proportion of individual facets, 2) quantification of volume fraction of pores, and their surface area, 3) quantification of Pt particles and their attachment to  $\text{Al}_2\text{O}_3$ . While this topic will be the focus of our future work, an essential qualitative assessment of the  $\gamma$ -Alumina surfaces and geometry of the Pt NPs can be obtained by transforming the stack of predicted segmentations into 3D visualizations. Figure 6a–d shows 3D volume visualization of  $\gamma$ -Alumina and Pt NPs from different viewpoints, and surface contour maps in



**Figure 6.** 3D visualization of  $\gamma$ -Alumina/Pt catalytic particle from different viewpoints. Pt NPs are colored in red, while  $\gamma$ -Alumina support material is colored in light gray. (a–c) Showing periodic facets on the {110} type broad surfaces of  $\gamma$ -Alumina along with Pt NPs. (d) Showing the pores inside the  $\gamma$ -Alumina matrix in a 3D transparent view. 3D visualizations were generated using open-source Paraview<sup>47</sup> software version 5.9.1 (<http://www.paraview.org>).

light gray represent  $\gamma$ -Alumina and red Pt NPs. 3D volume visualizations show that {110} surfaces of  $\gamma$ -Alumina are not atomically flat; instead, they form a series of periodically repeating structural facets. These facets are mostly terminated towards the center of  $\gamma$ -Alumina, and Pt NPs are anchored along with the {111} type facets rather than randomly distributed on the surfaces. Surprisingly, matrix pores are aligned along the direction of the surface facets, as seen in Fig. 6d.

We have demonstrated the effectiveness of a deep learning model in multi-class semantic segmentation of large and unbalanced data. This work was in many respects exploratory as a proof of concept, with a focus mainly on model performance. The current model is naturally limited in its applicability. Alumina-based catalysts with various morphologies and particle contrasts will require additional adaptations to the current model. We intend to continue to utilize this U-Net model in a transfer learning environment, incorporating the broad morphological and size variation of Alumina-based catalysts and eventually with general catalyst systems to achieve wider applicability. With recent advances in automatic data collection, deep learning-assisted semantic segmentation is genuinely expected to broaden the field of STEM tomography for routine quantitative measurement of catalysts on a statistically relevant scale, which is not possible currently.

## Methods

**3D tomography data acquisition and visualization.** Our main goal is to assess the effectiveness of a deep learning-based approach in semantic segmentation of the 3D HAADF STEM tomography reconstructions while achieving a full 3D view of the  $\gamma$ -Alumina/Pt catalytic material. For this purpose, we conducted HAADF STEM tomography experiments on a well-isolated  $\gamma$ -Alumina/Pt catalytic particle. TEM samples were prepared by dropping a solution containing well-dispersed NPs on a lacey carbon film. A detailed description of the synthesis of  $\gamma$ -Alumina/Pt material was reported in an earlier paper<sup>24</sup>. A probe aberration-corrected 300 kV Thermo Fisher Scientific Titan S/TEM microscope was used to acquire the HAADF STEM tilt series. HAADF STEM images were acquired at the detector inner collection angle of 40 mrad, beam current of 20 pA with a 0.1 nm probe size, an accelerating voltage of 200 kV. To extend the depth of focus of the electron beam during STEM tomography acquisitions, the convergence angle of the illumination system was adjusted to 10 mrad using the three-condenser lens optics of the microscope. Tomography tilt series consists of 69 HAADF STEM images



acquired at the tilt range of  $\pm 68^\circ$  and tilt increment of  $2^\circ$ . Post-processing of the tilt series was conducted using open-source resources; a Python programming script based on the tomviz software used for the image shift and tilt alignments<sup>44</sup>, and TomoPy and ASTRA Toolbox Python libraries for the maximum likelihood expectation maximization (MLEM) reconstructions<sup>45,46</sup>. Paraview software was employed to generate 3D visualizations from the fully segmented 3D reconstructions<sup>47</sup>.

**Training and optimization.** For optimization, we used a mini-batch gradient descent with a batch size of 2 and Adam optimizer at a learning rate of 0.0005. We used default parameters from the Tensorflow deep learning framework for the first and second moments of gradient averaging and updating<sup>48</sup>. All the weights were initialized by “He normal” kernel initialization<sup>49</sup>, and all the biases were initialized at 0. A total of 30 ground-truth images were selected from the 3D reconstructions, and corresponding ground-truth segmentations were manually annotated for the training and validation steps. A class label representing background/pores,  $\gamma$ -Alumina and Pt NPs, were assigned to each pixel in the ground-truth segmentations. Due to the large image size and limited GPU memory availability,  $1024 \times 512$  pixels (0.12 nm/pixel) ground-truth images and segmentations were divided into  $512 \times 512$  pixels patches. This data was then randomly split into 75% training and 25% validation data sets. The average pixel density of each class in the patches is 77.7% for background/pores, 21.8% for  $\gamma$ -Alumina, and 0.5% for Pt NPs. Data augmentation is crucial to teach the network a robust invariance to input data and generalize the model. We used rotation, vertical and horizontal flip, zoom, and shear transformations during training to generate a diverse range of images representing variations in the location and shape of the features. The best-performing model was selected based on the evaluation performance and applied to the automatic segmentation of a stack of 702 3D reconstructions. We employed a smooth blending approach to form final predictions where  $512 \times 512$  pixels size segmented patches were smoothly merged into  $1024 \times 512$  pixels size final predictions using spline interpolations between the overlapping patches<sup>50</sup>.

### Data availability

Data sets and Python code used is publicly available in the link provided below.

### Code availability

Python code for U-Net architecture, training and evaluation of the results are available at [https://github.com/ArdaGen/Multi\\_Class\\_Semantic\\_Segmentation.git](https://github.com/ArdaGen/Multi_Class_Semantic_Segmentation.git).

Received: 28 February 2022; Accepted: 11 July 2022

Published online: 28 September 2022

### References

- Arslan, I., Yates, T. J. V., Browning, N. D. & Midgley, P. A. Embedded nanostructures revealed in three dimensions. *Science* **309**, 2195–2198 (2005).
- Midgley, P. A. & Weyland, M. 3D electron microscopy in the physical sciences: The development of Z-contrast and EFTEM tomography. *Ultramicroscopy* **96**, 413–431 (2003).
- Midgley, P. A. *et al.* Nanoscale scanning transmission electron tomography. *J. Microsc.* **223**, 185–190 (2006).
- Midgley, P. A., Weyland, M., Thomas, J. M. & Johnson, B. F. G. Z-Contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering. *Chem. Commun.* 907–908. <https://doi.org/10.1039/b101819c> (2001).
- Weyland, M., Midgley, P. A. & Thomas, J. M. Electron tomography of nanoparticle catalysts on porous supports: A new technique based on Rutherford scattering. *J. Phys. Chem. B* **105**, 7882–7886 (2001).
- Gürsoy, D. *et al.* Rapid alignment of nanotomography data using joint iterative reconstruction and reprojection. *Sci. Rep.* **7**, 11818 (2017).
- Wang, C., Ding, G., Liu, Y. & Xin, H. L. 0.7 Å resolution electron tomography enabled by deep-learning-aided information recovery. *Adv. Intell. Syst.* **2**, 2000152 (2020).
- Han, Y. *et al.* Deep learning STEM-EDX tomography of nanocrystals. *Nat. Mach. Intell.* **3**, 267–274 (2021).
- Shepp, L. A. & Vardi, Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**, 113–122 (1982).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* Vol. 25 (eds Pereira, F. *et al.*) (Curran Associates, Inc., 2012).
- Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3431–3440 (IEEE, 2015). <https://doi.org/10.1109/CVPR.2015.7298965>
- Ciresan, D., Giusti, A., Gambardella, L. & Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems* Vol. 25 (eds Pereira, F. *et al.*) (Curran Associates, Inc., 2012).
- Milletari, F., Navab, N. & Ahmadi, S.-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. in *2016 Fourth International Conference on 3D Vision (3DV)* 565–571 (IEEE, 2016). <https://doi.org/10.1109/3DV.2016.79>
- DeCost, B. L., Lei, B., Francis, T. & Holm, E. A. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microsc. Microanal.* **25**, 21–29 (2019).
- Dong, H., Yang, G., Liu, F., Mo, Y. & Guo, Y. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. [arXiv:1705.03820](https://arxiv.org/abs/1705.03820) (2017).
- Greenspan, H., van Ginneken, B. & Summers, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging* **35**, 1153–1159 (2016).
- Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* Vol. 15 (eds Gordon, G. *et al.*) 315–323 (PMLR, 2011).
- Zeng, T., Wu, B. & Ji, S. DeepEM3D: Approaching human-level performance on 3D anisotropic EM image segmentation. *Bioinformatics* **33**, 2555–2562 (2017).
- Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01094-0> (2021).
- Roberts, G. *et al.* Deep learning for semantic segmentation of defects in advanced STEM images of steels. *Sci. Rep.* **9**, 12744 (2019).

22. Hernández-García, A. & König, P. Data augmentation instead of explicit regularization. [arXiv:1806.03852](https://arxiv.org/abs/1806.03852) (2018).
23. Leach, B. E. *Applied Industrial Catalysis* Vol. 1 (Academic Press, 1983).
24. Kovarik, L. *et al.* Tomography and high-resolution electron microscopy study of surfaces and porosity in a plate-like  $\gamma$ -Al<sub>2</sub>O<sub>3</sub>. *J. Phys. Chem. C* **117**, 179–186 (2013).
25. Khivantsev, K., Jaegers, N. R., Kwak, J., Szanyi, J. & Kovarik, L. Precise identification and characterization of catalytically active sites on the surface of  $\gamma$ -Alumina\*\*. *Angew. Chem.* **133**, 17663–17671 (2021).
26. Roiban, L. *et al.* 3D-TEM investigation of the nanostructure of a  $\delta$ -Al<sub>2</sub>O<sub>3</sub> catalyst support decorated with Pd nanoparticles. *Nanoscale* **4**, 946–954 (2012).
27. Epicier, T. *et al.* 2D & 3D in situ study of the calcination of Pd nanocatalysts supported on delta-Alumina in an environmental transmission electron microscope. *Catal. Today* **334**, 68–78 (2019).
28. Qin, R. *et al.* Weighted focal loss: An effective loss function to overcome unbalance problem of chest X-ray14. *IOP Conf. Ser. Mater. Sci. Eng.* **428**, 012022 (2018).
29. Novikov, A. A. *et al.* Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Trans. Med. Imaging* **37**, 1865–1876 (2018).
30. Sugino, T. *et al.* Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. *Healthcare* **9**, 938 (2021).
31. Yeung, M., Sala, E., Schönlieb, C.-B. & Rundo, L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **95**, 102026 (2021).
32. Jadon, S. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* 1–7 (2020).
33. Sudre, C. H., Li, W., Vercauteren, T. K. M., Ourselin, S. & Cardoso, M. J. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep learning in medical image analysis and multimodal learning for clinical decision support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, held in conjunction with MICCAI 2017 Quebec City*, 240–248 (2017).
34. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318–327 (2020).
35. Horwath, J. P., Zakharov, D. N., Mégret, 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).
36. Akers, S. *et al.* Rapid and flexible segmentation of electron microscopy data using few-shot machine learning. *npj Comput. Mater.* **7**, 187 (2021).
37. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. in *MICCAI* (2015).
38. Mavrin, A. *Focal loss*. <https://github.com/artemavrin/focal-loss> (2022).
39. Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **15**, 29 (2015).
40. Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863 (1993).
41. Maiseli, B. J. Hausdorff distance with outliers and noise resilience capabilities. *SN Comput. Sci.* **2**, 358 (2021).
42. Sim, D.-G., Kwon, O.-K. & Park, R.-H. Object matching algorithms using robust Hausdorff distance measures. *IEEE Trans. Image Process.* **8**, 425–429 (1999).
43. *Surface distance metrics*. <https://github.com/deepmind/surface-distance> (DeepMind, 2022).
44. Tomviz for tomographic visualization of nanoscale materials. <https://tomviz.org/>
45. Gürsoy, D., De Carlo, F., Xiao, X. & Jacobsen, C. TomoPy: A framework for the analysis of synchrotron tomographic data. *J. Synchrotron. Radiat.* **21**, 1188–1193 (2014).
46. Pelt, D. M. *et al.* Integration of TomoPy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data. *J. Synchrotron. Radiat.* **23**, 842–849 (2016).
47. Ahrens, J. P., Geveci, B. & Law, C. C. ParaView: An end-user tool for large-data visualization. in *The Visualization Handbook* (2005).
48. Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous distributed systems. [arXiv:1603.04467](https://arxiv.org/abs/1603.04467) (2016).
49. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. in *2015 IEEE International Conference on Computer Vision (ICCV)* 1026–1034. <https://doi.org/10.1109/ICCV.2015.123> (2015).
50. *Make smooth predictions by blending image patches, such as for image segmentation*. <https://github.com/Vooban/Smoothly-Blend-Image-Patches> (Vooban, 2022).

## Acknowledgements

The experimental work was conducted in the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830. LK was supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences.

## Author contributions

L.K. collected the HAADF STEM tomography tilt series. A.G. conducted the tomography tilt series alignment, reconstruction, and 3D visualization. A.G. planned and executed the computational experiments on the deep learning study. A.G., L.K. and H.L.F. equally contributed to the writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to L.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](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/>.

© Pacific Northwest National Laboratory 2022