Precise Size Determination of Supported Catalyst Nanoparticles via Generative AI and Scanning Transmission Electron Microscopy

Author(s):
Eliasson, Henrik; Lothian, Angus; Surin, Ivan; Mitchell, Sharon; Pérez-Ramírez, Javier; Erni, Rolf

Publication date:
2024

Permanent link:
https://doi.org/10.3929/ethz-b-000698670

Originally published in:
Small Methods, https://doi.org/10.1002/smtd.202401108

small methods
www.small-methods.com

Precise Size Determination of Supported Catalyst Nanoparticles via Generative AI and Scanning Transmission Electron Microscopy

Henrik Eliasson, Angus Lothian, Ivan Surin, Sharon Mitchell, Javier Pérez-Ramírez, and Rolf Erni*

Transmission electron microscopy (TEM) plays a crucial role in heterogeneous catalysis for assessing the size distribution of supported metal nanoparticles. Typically, nanoparticle size is quantified by measuring the diameter under the assumption of spherical geometry, a simplification that limits the precision needed for advancing synthesis-structure-performance relationships. Currently, there is a lack of techniques that can reliably extract more meaningful information from atomically resolved TEM images, like nuclearity or geometry. Here, cycle-consistent generative adversarial networks (CycleGANs) are explored to bridge experimental and simulated images, directly linking experimental observations with information from their underlying atomic structure. Using the versatile Pt/CeO_2 (Pt particles centered ≈2 nm) catalyst synthesized by impregnation, large datasets of experimental scanning transmission electron micrographs and physical image simulations are created to train a CycleGAN. A subsequent size-estimation network is developed to determine the nuclearity of imaged nanoparticles, providing plausible estimates for ≈70% of experimentally observed particles. This automatic approach enables precise size determination of supported nanoparticle-based catalysts overcoming crystal orientation limitations of conventional techniques, promising high accuracy with sufficient training data. Tools like this are envisioned to be of great use in designing and characterizing catalytic materials with improved atomic precision.

1. Introduction

The performance of nanoparticle-based heterogeneous catalysts can be significantly influenced by the size, shape, and corresponding support-interface geometry of the dispersed species.[1] This structure sensitivity necessitates the ability to precisely characterize such properties on the atomic scale in a statistically robust manner to understand and optimize catalytic materials. Transmission electron microscopy (TEM) has become indispensable for directly observing and analyzing the structural features of small supported nanoparticles. However, the conventional method of manually or automatically measuring particle diameters fails to capture the true complexity of nanoparticle structures, including their nuclearity and geometry. This is unsatisfactory because nanoparticles of different shapes can have the same diameter when projected in a transmission electron micrograph but exhibit vastly different reactivity due to variations in exposed facets, atomic coordination, and support interactions which determine the geometric and electronic properties of the active phase.[2]

Pt/CeO_2 catalysts, which are widely explored for diverse applications such as thermo- and electrocatalytic selective oxidations,[2–4] selective hydrogenation,[5,6] the water-gas shift reaction,[7] the oxygen reduction reaction,[8] alcohol reforming,[9] and hydrogenolysis,[10] and have been identified among the most promising practical systems for automotive treatment solutions, are a well-known example catalytic material where the dispersed Pt phase can adopt

H. Eliasson, R. Erni
Electron Microscopy Center
Empa – Swiss Federal Laboratories for Materials Science and Technology
Überlandstrasse 129, Dübendorf 8600, Switzerland
E-mail: henrik.eliasson@empa.ch

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smtd.202401108

DOI: 10.1002/smtd.202401108

A. Lothian
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
Linköping 581 83, Sweden

I. Surin, S. Mitchell, J. Pérez-Ramírez
Department of Chemistry and Applied Biosciences
Institute for Chemical and Bioengineering, ETH Zürich
Vladimir-Prelog-Weg 1, Zürich 8093, Switzerland

various 2D and 3D shapes depending on factors like the ceria support facet, Pt content, presence of other promoters, and the synthetic approach.[11,12] Structure sensitivity to the Pt size, geometry, and the related interface with the support, has been demonstrated in most of these applications,[13,14] highlighting the inadequacy of diameter-based categorization in predicting catalytic performance accurately.

To address this challenge, advanced image analysis techniques are required to extract deeper structural information from TEM images of supported nanoparticles. For most quantitative TEM applications, the high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image modality is used. The HAADF-STEM signal carries chemical information by a mass-thickness contrast and is particularly appreciated due to the incoherent nature of the signal and ease of interpretation of the resulting images.[15–17] One quantitative application of HAADF-STEM is atom counting, where the number of atoms of an imaged nanostructure can be estimated by analyzing the pixel intensities of the atomic columns.[18–20] Atom counting has been explored by both statistical approaches and machine learning and can generally yield a size estimate with high precision.[19,21] However, it relies on the atomic columns of the region of interest being in a zone axis, which greatly limits the applicability of the technique to nanoparticles imaged under specific conditions. This emphasizes the need for a technique that can handle nanoparticles imaged from any orientation.

Recent advances have demonstrated the utility of multislice image simulations in directly comparing them with HAADF images to infer detailed atomic structures.[20,22–24] A multislice image is generated by simulating the propagation of the incident electron wave through an atomic model and integrating the scattered intensity over a virtual detector. The atomic model and corresponding multislice image simulation are thus paired, and it is possible to use supervised learning techniques like a U-Net,[25] to infer information about the atomic model (e.g., nuclearity, support-interface, and crystal structure) from the simulated images regardless of particle orientation. By establishing a consistent mapping between experimental images and simulated images, it should by extension also be possible to infer information about the underlying atomic structure of an experimental image.

The cycle-consistent generative adversarial network (CycleGAN) is a type of machine learning model developed to learn a mapping between two unpaired image domains.[26] While initial applications included mapping horses to zebras and photographs to paintings, the CycleGAN, and inspired networks have become more commonly adopted in scientific imaging in the last few years.[27–29] A CycleGAN has previously been applied to HAADF-STEM data to generate realistic STEM images of 2D materials.[30] These images were then used to train a defect detection network, which generalized very well to real experimental data. However, the CycleGAN remains unproven on HAADF-STEM images of more common 3D objects like nanoparticles, where a larger variety in the training data is needed and regularization techniques are necessary to stabilize model training due to the larger structural variation and effects of dynamical diffraction in the images.

In this work, we train a CycleGAN to map between multislice images of platinum nanoparticles supported on cerium dioxide and experimental high-resolution HAADF-STEM images of the same system. We use the learned mapping to generate synthetic experimental data with a known atomic model ground truth. These paired datasets are then utilized to train a size-estimator network in a supervised manner to predict the number of platinum atoms in experimentally observed platinum nanoparticles (**Figure 1**).

2. Results and Discussion

2.1. Generating Training Data

It is essential that both the simulated dataset and the experimental dataset are well balanced, meaning that they cover largely the same type of structures and are generated with similar imaging parameters for the sought mapping between them to retain physical relevance. The major imaging parameters like acceleration voltage, beam convergence angle, detector collection range, and effective source size were therefore kept similar. The acceleration voltage was 300 kV, the beam convergence angle 18.5 mrad, and the effective source size 50 pm for both the simulated and experimental data. The HAADF collection range was 66–140 mrad for the simulated data and 72–200 mrad for the experimental data. The reason for the discrepancy is explained in the section on simulated data.

While conventional STEM images typically depict the object of interest in a 'top-down' perspective,[31] for this project, we have chosen to image particles 'edge-on' or 'in profile' (as common in atom counting[32]). This approach was selected for several reasons. First, imaging the particles at the edge of the support material ensures minimal obstruction and little to no addition to the signal intensity, providing more information about the particle's shape by visualizing the wetting angle. Second, it allows us to maintain a small supercell as we are imaging thin edge regions. The heavy computational load of multislice simulations limits the size of the supercell. In this work, the typical supercell was $7 \times 7 \times 7$ nm or smaller, which is sufficient to cover a large variety of structures featuring nanoparticles at the edge of a support structure, but not in thicker support regions where many particles imaged top-down would be situated, potentially overlapping and suffering from an unknown top-bottom effect. An obvious drawback to this way of imaging is that data acquisition of a statistically significant number of particles is slower than when recording larger overview images, as many particles are not situated at a projected support edge. Although slower, we don't believe this is a significant limitation and also envision this method to be extendable to automatic data acquisition. One can imagine tracing the outline of agglomerates and recording images of particles along the border of the support as a feasible approach for automation.

2.1.1. Simulated Data

An algorithm to generate random particle structures on supports was developed to build a large simulated dataset of atomic models. A detailed explanation of the algorithm is available in the Supporting Information. In general, the workflow can be summarized in seven steps as visualized in **Figure 2**. First, the outline

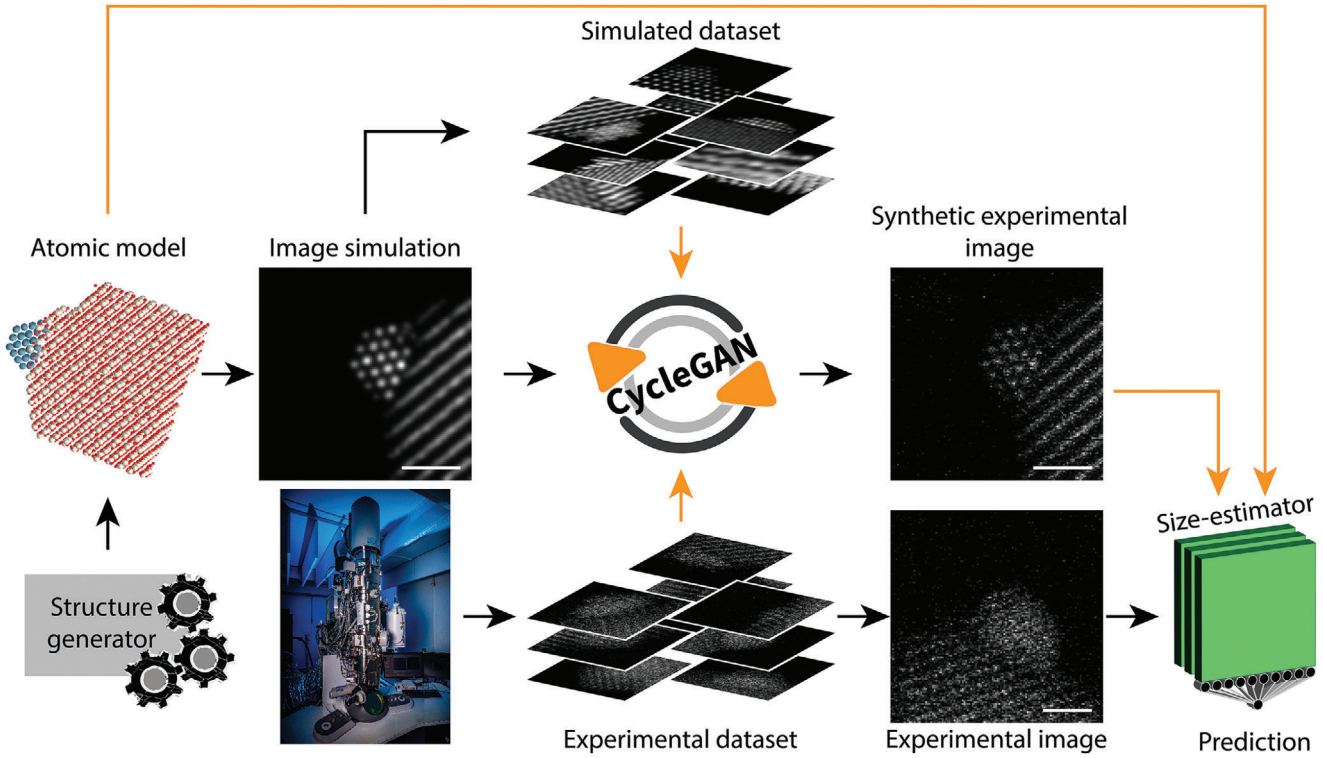

Figure 1. Overview of the approach: orange arrows indicate the data used to train the models and black arrows show the workflow. Atomic models are generated using the algorithm presented herein and are used to simulate a dataset of multislice images of Pt nanoparticles supported on ceria. An experimental high-resolution STEM dataset of nanoparticles in a Pt/CeO_2 catalyst is recorded with a Titan Themis microscope operated at 300 kV and used as training data for the CycleGAN together with the simulated dataset. The simulated dataset is mapped to the experimental domain using the trained CycleGAN, forming a synthetic dataset of "experimental-looking" images with corresponding atomic model ground truths. The synthetic dataset and the atomic models are used to train a subsequent size-estimator network. Once trained, the experimentally observed nanoparticles can be sent through the size-estimator, automatically outputting a particle size distribution in terms of the nuclearity (number of atoms present). The CycleGAN can also be used to denoise experimental data by sending experimental images through it in the opposite direction, yielding a denoised version of the image in the style of the multislice image simulations (see Figure 3). Each scale bar corresponds to 1 nm.

of the structure is generated, which is essentially a set of convex hulls stacked on top of each other, some of which house the nanoparticle while the others house the support material. A step edge can be added to the finalized structure, and the desired bulk lattices are then filtered within the outline of the particle and support regions, yielding an atomic model. Since the generation of the structure outline does not depend on a specific lattice, the algorithm could be used for any combination of elements and crystal structures. A subsection of the atomic model is then cropped and used as input for the image simulation.

Image calculations were carried out in Dr. Probe[24] and a script was created to automate the process such that no human input was needed to load the next structure. Each structure loaded into Dr. Probe included a buffer region of typically 1.5 nm around the actual object of interest to avoid periodicity artifacts in the generated images. All calculations used a slice thickness of 0.1 nm, a maximum spatial frequency k_{max} of 70 nm^{-1}, and 10 frozen lattice configurations per slice.[24,33] This k_{max} was chosen to keep computation time reasonably low while maintaining calculation accuracy. Increasing k_{max} further increased computation time significantly while yielding nearly identical images. However, as a consequence of the relatively low k_{max}, the collection range of the virtual HAADF detector is smaller than the one in the experimental setup.

In total, 5000 images were simulated, all with an image size of 128×128, a pixel size in the range of 5–35 pm, and featuring one Pt nanoparticle per image with a size in the range of 1–765 atoms. The mean computation time for an image was ≈10 min (with GPU acceleration) which is inconvenient and future work should work to establish a framework for large-scale parallel image simulations on supercomputers, perhaps utilizing newer image simulation algorithms like PRISM.[34,35] A corresponding noisy image was generated for each simulated image by adding Gaussian noise followed by Poisson noise, yielding a second "noisy" dataset. This noisy dataset was used along with the clean set to pretrain the CycleGAN. The manually added noise is simplified compared to real experimental noise but sufficient for pretraining where the goal is only to teach the CycleGAN general noising and denoising traits to ease model training with real experimental data. The reason why training with simulated images and the corresponding manually noised images is easier than training with simulated images and real experimental data is because the datasets are perfectly balanced in terms of the atomic structures behind the images.

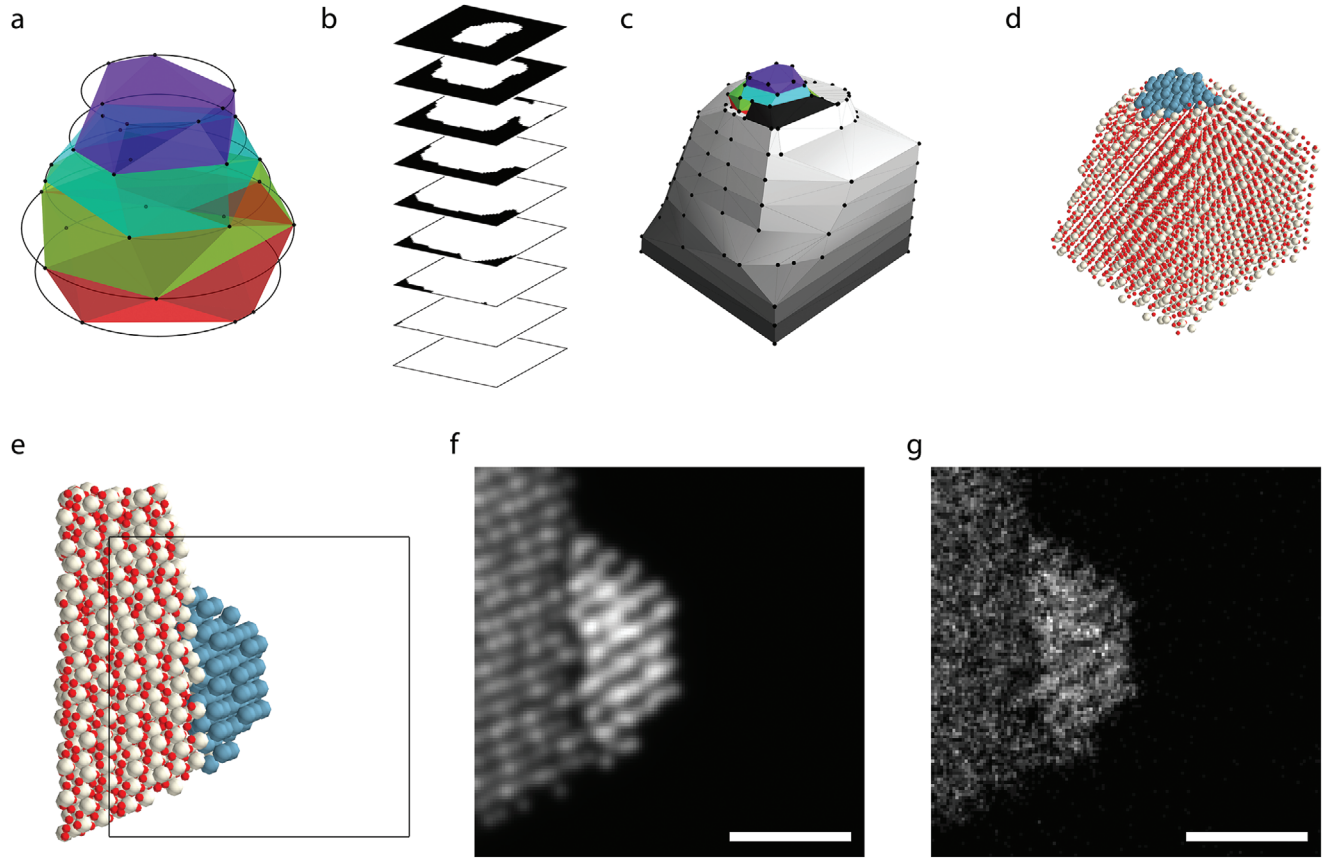

Figure 2. Key stages from the simulated data generation workflow illustrated for Pt/CeO_2. a) The metal nanoparticle structure outline. b) The nanoparticle footprint expanded by Perlin noise forming the outline of the support structure. c) The final structure outline with an added step-edge (black). d) The Pt nanoparticle and CeO_2 support crystal structures filtered within the structure outline. e) Cropped structure ready for input to the multislice simulation. f) Multislice HAADF image of the framed region in (e). g) The multislice image with Gaussian noise added and Poisson noise applied. Each scale bar corresponds to 1 nm.

2.1.2. Experimental Data

To efficiently record a significant number of experimental images while maintaining a manageable manual workload, we recorded time-series of each particle, leveraging natural noise variation and image drift to our advantage to increase variety in the experimental image set. Opting for time-series rather than longer exposure images also has the benefit of reducing electron beam-induced restructuring which is critical when imaging small supported nanoparticles since they often exhibit dynamic behavior. To maintain a consistent noise profile through all recorded images, the imaging parameters were kept constant with the exception of pixel size resulting from the use of different magnifications (see Experimental Section).

In total, 205 different nanoparticles from a Pt/CeO_2 catalyst were imaged, with some particles recorded at multiple magnifications, resulting in 360 time-series and a total of 11219 images. All imaged nanoparticles had a diameter in the range of 0–4 nm and were imaged with a pixel size in the range of 18–51 pm with the majority ≈25 pm like in the simulated dataset (Figure S1, Supporting Information).

2.2. Training the CycleGAN

The CycleGAN comprises four networks trained together in an adversarial manner: two generators and two discriminators. The generators are trained to learn the mapping functions $\boldsymbol{G}$: $X \rightarrow Y$ and $\boldsymbol{F}$: $Y \rightarrow X$, with X being the domain of the multislice image simulations, and Y being the experimental (or noisy) domain. Simultaneously, the discriminators are trained to discriminate between images generated by the generators and real images, forcing the generators to generate increasingly realistic images. However, the CycleGAN is notoriously difficult to train, often suffering from instabilities and mode collapse. Much effort has been dedicated to stabilizing model training through architectural modifications. One common issue is that the discriminators tend to overpower the generators. To counter this, techniques such as gradient penalty,[36] spectral normalization,[37] and the addition of an image buffer[38] to feed the discriminators with historically generated images have been found helpful. Additionally, for the generator architecture, replacing batch normalization layers with instance normalization has been shown to improve the quality of generated images in style transfer tasks such as mapping between two noise profiles.[39]

Our generators use a 4-layer deep U-Net architecture with instance normalization, while the discriminators employ the standard PatchGAN architecture with spectral normalization.[26,40] We experimented with smaller U-Net generators and ResNet[41] generators but found a 4-layer deep U-Net to be a good trade-off in model complexity, memory allocation, training time, and output quality. The models are available on GitHub: (https://github.com/benke97/PtCeriaGAN).

The generator loss function follows a standard form, incorporating three terms: a binary cross-entropy adversarial loss term ($\mathcal{L}_{adv,G}$), a cycle-consistency loss term ($\mathcal{L}_{cyc}$), and an identity loss term ($\mathcal{L}_{id}$). The overall loss function is expressed as:

$$\mathcal{L}_G = \mathcal{L}_{adv,G} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{id}\mathcal{L}_{id} \quad (1)$$

where λ_{cyc} and λ_{id} are tunable scaling parameters. The adversarial loss for the generator, $\mathcal{L}_{adv,G}$, encourages the generator to produce images that the discriminator cannot distinguish from real images. The cycle-consistency term ensures that the learned mappings are reversible, i.e., $\boldsymbol{F}(\boldsymbol{G}(x)) \approx x$ and $\boldsymbol{G}(\boldsymbol{F}(y)) \approx y$, $x \in X$, $y \in Y$. The identity loss ensures that images from one domain remain unchanged when sent through the generator that maps to their domain, meaning $\boldsymbol{F}(x) \approx x$ and $\boldsymbol{G}(y) \approx y$.

For comparisons between images in the simulated domain, we use an L1 loss for $\mathcal{L}_{cyc}$ and $\mathcal{L}_{id}$, while a perceptual loss calculated with LPIPS[42] using SqueezeNet[43] is employed for image comparisons within the experimental and noisy domains. The reasoning behind the use of perceptual loss for the noisy domain is that the noise in the input image is inherently random, making it counterproductive to recreate the exact pixel values as long as the overall noise profile remains consistent. For a mathematical formulation of the loss function, see other works.[26] For a visual description of the losses, see Figures S2–S4 (Supporting Information).

The adversarial loss for the discriminator, $\mathcal{L}_{adv,D}$, is a binary cross-entropy loss that encourages the discriminator to classify generated images as fake and real images as real, thus opposing the generator's objective. To stabilize training, we also experimented with a gradient penalty term ($\mathcal{L}_{GP}$) which was added to the discriminator loss, resulting in the full expression of the discriminator loss:

$$\mathcal{L}_D = \mathcal{L}_{adv,D} \left(+ \lambda_{GP}\mathcal{L}_{GP}\right) \quad (2)$$

where λ_{GP} is another scale factor. In the end, we found that spectral normalization (with the default parameters of the PyTorch implementation) along with an image buffer resulted in perceptually higher quality images compared to when gradient penalty was used along with the image buffer. Using gradient penalty together with spectral normalization inhibited the creativity of the generators too much in our case but could be worth exploring for applications with more limited datasets.

The CycleGAN was pretrained using the clean and noisy simulated datasets. The sets were min-max normalized within their domain and split into training and validation sets (4:1), 4000 training images, and 1000 validation images. The optimal pretrained weights were determined by evaluating the reconstruction quality. This quality was measured using the mean squared error between the simulated images and the denoised versions of their paired noisy images, along with the perceptual loss between the original noisy images and the noised version of their paired simulated images. We pretrained the model for 100 epochs with the Adam optimizer ($(\beta_1, \beta_2) = (0.5, 0.999)$) and constant learning rates of: lr_G = 5e-5 and lr_D = 7.5e-5. Furthermore, $\lambda_{cyc} = 5$, $\lambda_{id} = 0.5\lambda_{cyc}$, $\lambda_{GP} = 0.0001$, a batch size of 1, and an image buffer of 20 images for each discriminator was used. Both generators were trained with the same optimizer for simplicity.

The pretrained generators were used as a starting point for the main task of learning the mapping functions $\boldsymbol{F}$ and $\boldsymbol{G}$ between the dataset of multislice simulations and the experimental dataset. A subset of 5000 experimental images was selected as the experimental dataset, sampling images from all 360 time series. This experimental set was split into training and validation sets (4:1), ensuring that none of the 205 nanoparticles appear in both training and validation sets. Also here, the simulated and experimental sets were min-max normalized within their own domain.

We train the CycleGAN for 200 epochs, again with the Adam optimizer ($(\beta_1, \beta_2) = (0.5, 0.999)$) using a constant learning rate of 1e-5 for all networks and with the three bottom layers of the generator encoders frozen. From quick tests, it is likely that freezing these layers had little effect on the outcome. Other hyperparamters used were: $\lambda_{cyc} = 300$, $\lambda_{id} = 0.25\lambda_{cyc}$, and $\lambda_{GP} = 0$, batch size = 4, and an image buffer of 20 images for each discriminator. The generators were trained with the same optimizer for simplicity. The need for the high λ_{cyc} stems from the normalization of the datasets. Due to the large variation in pixel intensities between some images in the simulated set and the different intensity distributions in the experimental and simulated sets, a large λ_{cyc} was needed to bring the cycle-consistency loss term to the same scale as the adversarial loss term. Using a smaller λ_{cyc} caused the generators to introduce more noise into the generated experimental images to fit the distribution of the real experimental set. This made the images overly noisy and less realistic with respect to the underlying structure, despite appearing realistic. Consequently, the denoising process began to discard some structural features such as noise, which is undesirable for physical denoising.

The best model weights were determined by tracking the Fréchet Inception Distance (FID), a commonly used metric to determine the quality of images generated by a GAN, between the generated image sets and their respective real set during training.[44] The epoch with the lowest sum of the FID scores for the two domains was selected. A selection of generated images from the best model is displayed in **Figure 3**. For the experimental domain, we achieve a FID score of 59.7 and for the simulated domain we achieve a FID score of 94.3. These values are fairly high and by using a lower λ_{cyc} we could achieve a FID score below 30 for the experimental domain, but at the unacceptable trade-off in physical realism described previously. This hints rather at the FID score punishing the difference in the underlying imaged objects which is reasonable as our atomic models are too sterile and rigid to cover the complexity of the experimentally observed structures and their dynamics during STEM acquisition.

By manual inspection, the $\boldsymbol{G}$ mapping generates realistic experimental images with a noise profile that closely resembles that of the experimental data. There are however some fail cases like for images with a pixel size smaller than ≈10 pm where the generated images do not look convincing. In certain cases, mainly

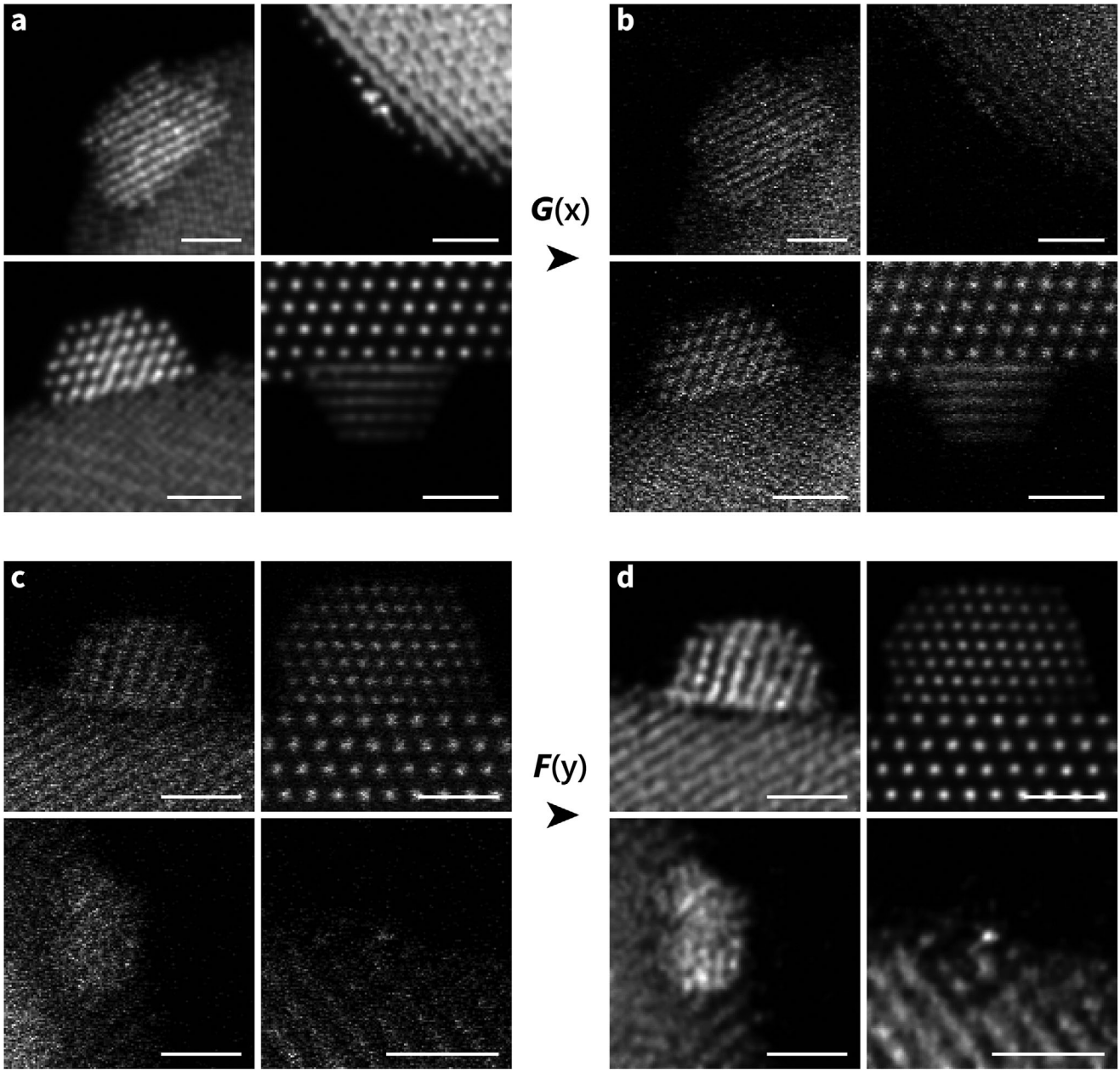

Figure 3. A selection of Pt/CeO_2 images from a) the simulated dataset, b) the synthetic experimental dataset, c) the experimental HAADF-STEM dataset, and d) the denoised experimental dataset. Each scale bar corresponds to 1 nm.

images with low pixel intensities, the speckle noise in the vacuum background can also seem overly pronounced and there are instances of black cavities generated in the middle of the support structure.

The denoising mapping $\boldsymbol{F}$ works well despite the high FID score, maintaining the intensity distribution of the input image while removing noise and not introducing any apparent artifacts. Lattice features visible in the experimental images are, by visual inspection of a subset of images, always retained in the denoised image. Of course, the true denoised state of an experimental image is subject to discussion, and the experimental noise likely makes certain information unrecoverable.

Although the denoising generator is specialized on a specific noise profile, it is apparent that it is somewhat robust and generalizes decently to atomically resolved time-series of particles recorded under different imaging conditions (Video S1, Supporting Information). One of the investigated time-series was recorded in situ with a gas-cell holder in 10% H_2/90% N_2 at 1000 mbar and 300 °C. Currently, atomically resolved data from gas-cell holders suffers from significantly reduced temporal resolution due to the encapsulating SiN windows and the gas atmosphere, making it difficult to study the dynamics and fluxional behavior of supported nanoparticles under reaction conditions.[45] This in situ series, recorded at 5 frames s^{-1}, was denoised by our

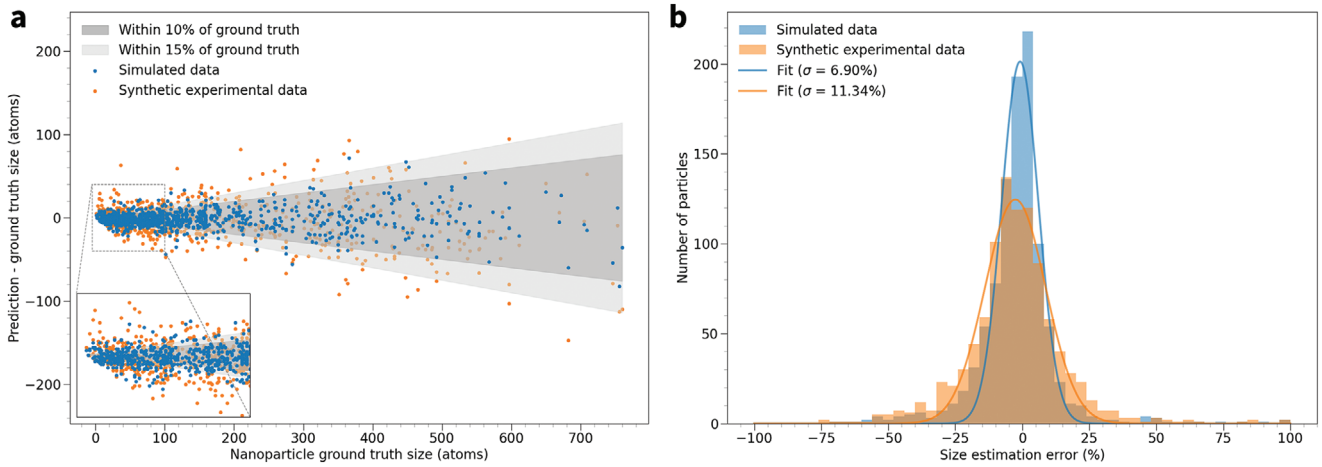

Figure 4. a) The nanoparticle size predictions subtracted by ground truth size for the 1000 Pt particles in the validation images, evaluated both in the simulated domain and as synthetic experimental data. b) The typical size estimation error in percentage reveals a Gaussian distribution of the errors where the standard error in the simulated domain is 6.90% and 11.34% in the experimental domain.

generator after fine-tuning for 1 iteration on a small dataset of Pt particles recorded in situ. Although not perfect, this highlights a potential for the CycleGAN to aid in improving the temporal resolution of in situ gas-cell data.

2.3. Particle Size Estimation

With the learned mappings G and F, we now have two options for training a size-estimator network to generalize to the experimental data. One option is to train the size-estimator directly on the multislice simulations and use the denoised experimental data as test data. Alternatively, we can train the network on synthetic experimental data and use the raw experimental data as test data. To compare the two approaches, we trained two size-estimators: one on synthetic experimental data and the other directly on the simulated images. We then evaluated their performance in predicting the size of nanoparticles in terms of number of atoms.

We use a U-Net with 4 down-sampling steps, dropout, and batch normalization for the size-estimator network. Global average pooling followed by a linear layer was added to the end of the decoder to output a single value. The U-Net architecture was chosen for its effectiveness in handling tasks that require multilevel feature extraction. Estimating a particle size from a STEM image requires multiple levels of understanding, the network needs to identify the particle within the image and differentiate it from the support and vacuum regions, detect subtle lattice features to infer pixel size (magnification), and evaluate shape and pixel intensities to estimate the actual size. The U-Net's ability to combine detailed high-resolution features with abstract, high-level information via skip connections makes it suitable for this task.

The 5000 simulated images were split into training and validation sets (4:1), and synthetic experimental data was generated for the experimental domain size-estimator by sending them through the G generator. Both size-estimators were trained for 250 epochs using an L1 loss function, a batch size of 32, and the Adam optimizer with a cosine annealing learning rate scheduler with an initial learning rate of 0.0008. We also explored using MAPE loss, MSLE loss, and MSE loss but the L1 loss was chosen as it yielded a slightly higher fraction of predictions within a satisfactory error margin (15%) of the ground truth size. The predictions of the 1000 particles in the validation set and their errors are presented in **Figure 4**. Of the predictions outputted by the size-estimator trained on the simulated set, 72.6% are found within 10% of the ground truth size. For the experimental domain size estimator, the predictions have a larger mean error and 70.1% of predictions are within 15% of the target. Fitting a normal distribution to the histogram of the errors reveals a standard error of ≈6.9% for the simulated domain and 11.3% for the experimental domain.

The performance on simulated data is within our expectations considering the relatively complex task the size-estimator is faced with and the still relatively small dataset of 5000 images. While 5000 simulated STEM images could be considered large in the TEM community, to put the dataset size into perspective, the KITTI dataset contains over 90 000 images for the task of depth evaluation in autonomous driving scenarios.[46] It is possible that more modern architectures like the vision transformer would make for a better size-estimator architecture when more data is available.[47]

The trained size-estimators were applied to 205 experimentally observed nanoparticles. To verify the generalization of the model, all predictions were manually inspected and compared to the conventional hemisphere estimate technique where the particle diameter is manually measured and the particle is assumed to have a hemispherical shape and a Pt atom density of an ideal Pt fcc lattice. A prediction was deemed plausible based on visual inspection, keeping in mind that the hemisphere estimate is typically an underestimate for larger particles. This process is of course subject to human bias, but because there is no ground truth available for the current experimental data, human evaluation is unavoidable.

For each nanoparticle, predictions were averaged across all frames within the recorded time-series. We thus assume the particle maintains its size throughout the series which was deemed reasonable by visual inspection. When multiple time-series were

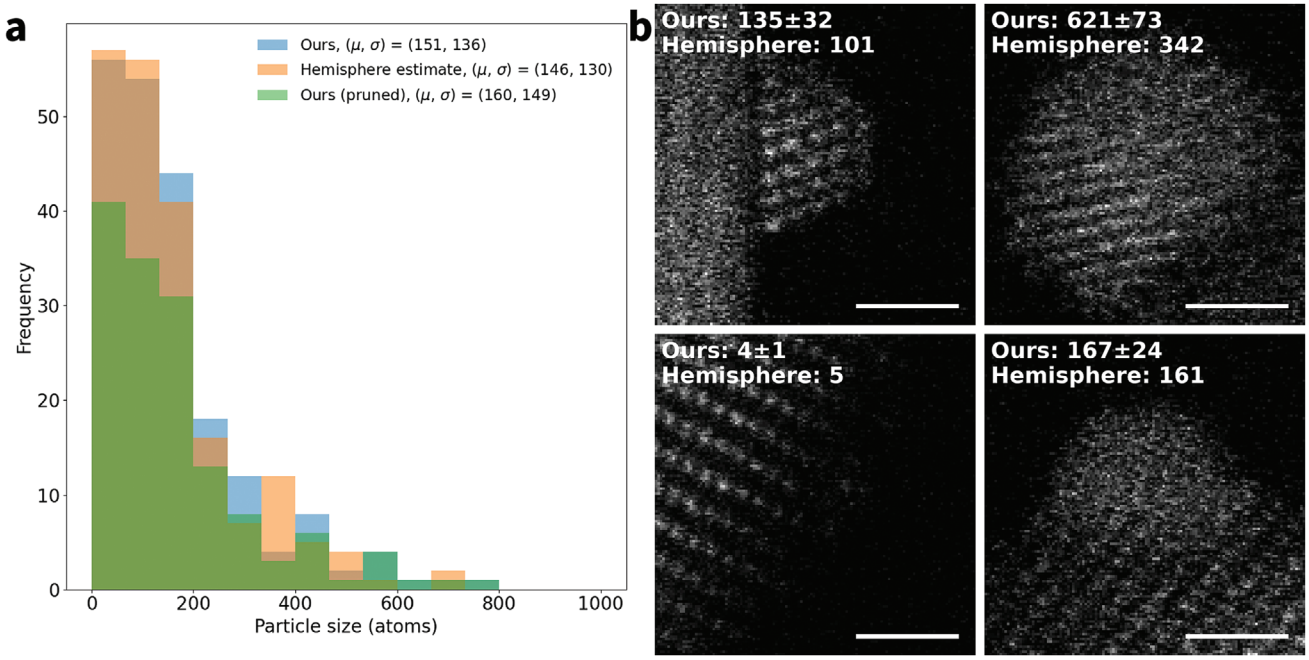

Figure 5. a) Particle size distribution of 205 Pt nanoparticles in the Pt/CeO_2 catalyst. b) A selection of experimental particles with size estimates by our model and by manual diameter assuming a hemispherical particle shape and an ideal Pt fcc crystal. The image showing a small cluster predicted to be 4–5 atoms is displayed with higher contrast in Figure S9 (Supporting Information). The model handles particles of widely varying atom counts in the range 1–1000. Each scale bar corresponds to 1 nm. A collage of all the 205 predictions can be found in the Supporting Information.

available for a single particle, only the longest series with a pixel size closest to 25 pm was used for the final estimation. The error of a prediction in number of atoms was calculated by:

$$\epsilon = \sqrt{(\mu_t \sigma_{model})^2 + (\sigma_t)^2} \quad (3)$$

where σ_t and μ_t are the standard deviation and mean of all predictions within a time-series in the number of atoms, and σ_{model} is the standard error of the size-estimator in percent. We used $\sigma_{model} = 6.90\%$ for the simulated domain and $\sigma_{model} = 11.34\%$ for the experimental domain.

We found that the size-estimator trained on synthetic experimental data and applied directly to the raw experimental data outputted a plausible estimate for 70.7% of the particles (see the Supporting Information for a collage of all 205 predictions). This was slightly higher than for the size-estimator trained on simulated data and applied to denoised experimental data where 69.8% of the predictions were deemed plausible. This small difference indicates that the choice of working domain is not that significant. Some plausible estimates by the experimental domain size-estimator are displayed in **Figure 5** and some fail-cases are presented in Figure S5 (Supporting Information). There is no clear trend related to measured diameter or pixel size explaining for what type of images the network fails (Figure S6, Supporting Information). We believe the model's performance is mainly limited by data shortages and the complexity of including a range of pixel sizes in the datasets. Other reasons may include poor experimental image quality (incl. instrument instabilities), large changes in the image and particle structure during a time-series, e.g. introduction of channeling contrast when a particle rotates in and out of a zone axis, and unrepresentative particle structures in the simulated set.

Looking at the particle size distribution, we see that our model yields a particle size distribution similar to that of the manual hemisphere estimation technique, but with a slightly larger mean particle size (Figure 5) and a smaller standard deviation (Figure S8, Supporting Information) after pruning the unrealistic predictions. The size-estimator trained directly on simulated data yields a particle size distribution with a slightly higher mean and standard deviation than its counterpart trained on synthetic experimental data, both before and after pruning the unrealistic predictions (Figures S7 and S8, Supporting Information).

3. Practical Considerations

The models presented herein are trained specifically for Pt/CeO_2 and the specific noise profile of our Titan Themis microscope with a fixed set of imaging parameters. To apply the methods to other metals or supports, both a new simulated dataset and an experimental dataset would have to be generated, and the models retrained. The structure generator could still be reused to generate new atomic structures, as it is agnostic to the element and crystal structures used. No matter the system, our trained CycleGAN weights could likely be used as a starting point for fine-tuning the model to other systems with new datasets.

To study Pt/CeO_2 with another microscope but with similar imaging parameters and particle size range, as we have, only a new experimental dataset would have to be recorded as our simulated dataset could be reused. Although we are only concerned with HAADF-STEM images in this work, bright field (0–10 mrad)

and annular dark field (10–24 mrad) STEM images are also available for each of the 5000 simulated structures.

Given the image size of 128 × 128 pixels, the pixel size will determine how large a particle can fit within the frame, effectively limiting the applicable size range of the technique, e.g. up to ≈4 nm for a pixel size of 25 pm. The networks could be retrained for larger image sizes but that would require larger supercells for the image simulation, which is computationally expensive.

The technique is extendable to operando measurements with gas-cell holders, with the main difference being the new noise profile stemming from the encapsulating SiN windows and the gas environment. However, the acquisition of experimental data in a gas cell is significantly more demanding than corresponding measurements in a vacuum, suffering from a reduced spatial and temporal resolution, limited holder tilt range, and gas flow-related instabilities, which will make recording a large enough dataset more time-consuming. Since the properties of supported nanoparticles may vary fluxionally during operation depending on the conditions,[48,49] integrating these methods into in situ microscopy pipelines will be crucial.

Furthermore, the workflow can be adapted to other tasks than size estimation. As long as labels are available for the simulated atomic models, virtually any descriptor could be predicted. Our simulated dataset however, does not contain any labels for descriptors such as geometry, crystallinity, or particle-support interface, nor does our structure generator allow such control of the generated structures in its current state.

The limiting factor of this methodology is currently the slow process of generating simulated and experimental data. To address this, a framework for large-scale STEM image simulations on supercomputers alongside automated experimental image acquisition schemes should be implemented. With more data available, an improved descriptor-estimator architecture could be developed to explore the information limit of STEM images, potentially utilizing modern vision transformers. This would enhance our ability to characterize catalytic materials with unprecedented precision, driving advances in heterogeneous catalyst design.

4. Conclusion

We have presented a method that utilizes the CycleGAN model to effectively map between high-resolution HAADF-STEM images and physical multislice image simulations, enabling inference of atomic structural information from experimentally observed 3D configurations. We have shown this by training a size estimator that could output reasonable size predictions for 70.7% of 205 observed nanoparticles in a Pt/CeO_2 catalyst. With more data and by focusing on a single STEM magnification, significant further improvements in accuracy are foreseeable. Additionally, we found no significant difference in the generalization ability of the size-estimator network whether trained on clean simulated data and applied to denoised experimental data, or trained on synthetic CycleGAN-generated experimental data and applied directly to raw experimental data.

What sets this technique apart from conventional atom counting is that it is fully automatic and independent from zone axis constraints during imaging, making it more broadly applicable, especially for statistical analyses. Additionally, compared to manual size estimates and shape assumptions, this technique has the potential to extract as much information as is physically possible from STEM images when sufficient data is available.

The networks can be readily retrained for other supported nanoparticle-based catalysts beyond Pt/CeO_2 and for other imaging conditions including operando measurements, or to predict other atomic-scale descriptors beyond nuclearity by generating new data. Tools to extract atomic level descriptors will become increasingly valuable as researchers focus more on the atomic structure of materials, particularly in the precision design of emerging low-nuclearity heterogeneous catalysts or more complex ternary systems. The insights gained from our model training will significantly accelerate the application of CycleGANs for these purposes.

5. Experimental Section

Machine Learning and Data Analysis: All data analysis as well as development and training of the neural networks was done on a workstation equipped with an Intel Core i9-13900K Processor, an RTX4090 GPU, and 64 GB of RAM, using python and pyTorch. One training iteration for the CycleGAN took ≈1 min.

Scanning Transmission Electron Microscopy: Experimental STEM data was recorded on a probe-corrected Titan Themis S/TEM microscope operated at 300 kV. Samples were prepared by dispersing the Pt/CeO_2 powder in methanol and sonicating the solution for a few seconds followed by drop casting onto a standard Cu TEM grid covered by a lacey carbon film. All data was recorded in 512 × 512 time-series with a beam current of 35–40 pA, a beam dwell time of 1 µs, a convergence angle of 18.5 mrad, a HAADF collection angle from 72 to 200 mrad, a Gain of 43.16 dB, an Offset of −1.8, and a pixel size in the range 18–51 pm. For all recorded time-series, a 128 × 128 window around the particle of interest was cropped out, forming the training data. In some cases the time-series were rigidly aligned before cropping out the particle, in other cases, the drift was left in to increase the variability of the training data. In situ, gas-cell data was recorded using the DENSsolutions Climate GVB system.

Catalyst Synthesis: Pt/CeO_2 catalyst was prepared by adapting the procedure described elsewhere.[50] Namely, 21 mg of $H_2PtCl_6{\cdot}6H_2O$ (abcr, 99% purity metal basis) and 32 mg of NaOH (VWR chemicals, reagent grade) were each dissolved in 2 cm^3 of ethylene glycol (Acros Organics, 99.9%) to yield 0.04 and 0.4 M solutions, respectively. The solutions were combined in a glass vial and heated for 3 min at 413 K in a microwave reactor (CEM discovery, CEM Corporation, USA) applying a power of 100 W. Subsequently, 30 cm^3 of 1 M HCl (VWR chemicals, reagent grade) was added to the particle suspension, and the mixture was centrifuged at 6000 rpm (4400 rcf) for 5 min. The liquid was decanted and the washing/centrifugation procedure was repeated 3 times. The particles were then suspended in acetone (Sigma–Aldrich, 99.5%), 800 mg of CeO_2 powder (Sigma–Aldrich, nanopowder < 25 nm) was added and the mixture was sonicated in an ultrasonic bath for 1 h until acetone completely evaporated. The solid was dried overnight under a vacuum at 343 K.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors would like to thank Juri Barthel for assistance with Dr. Probe. H.E. and R.E. acknowledge funding from the Swiss National Science Foundation (200021_196381).

Open access funding provided by ETH-Bereich Forschungsanstalten.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

H.E. acquired the experimental data and generated the simulated data. H.E. devised, trained, and evaluated the neural networks in consultation with A.L. The Pt/CeO_2 catalyst was synthesized by I.S. The work was supervised by R.E., S.M., and J.P.-R. All authors have read, approved, and contributed to the manuscript.

Data Availability Statement

Model weights and data used to train and evaluate the models can be found on Zenodo: https://doi.org/10.5281/zenodo.12623013.

Keywords

ceria, cyclegan, generative AI, heterogeneous catalysis, nanoparticles, platinum, scanning transmission electron microscopy

Received: July 19, 2024
Revised: September 4, 2024
Published online:

[1] C. Vogt, B. M. Weckhuysen, *Nat. Rev. Chem.* **2022**, *6*, 89.
[2] B. Gao, K. Zhang, Y. Wang, J. Lu, P. Liu, *ChemCatChem* **2023**, *15*, 202301160.
[3] A. M. Gänzler, M. Casapu, F. Maurer, H. Störmer, D. Gerthsen, G. Ferré, P. Vernoux, B. Bornmann, R. Frahm, V. Murzin, M. Nachtegaal, M. Votsmeier, J.-D. Grunwaldt, *ACS Catal.* **2018**, *8*, 4800.
[4] Y. Lykhach, A. Figueroba, T. Skála, T. Duchoň, N. Tsud, M. Aulická, A. Neitzel, K. Veltruská, K. C. Prince, V. Matolín, K. M. Neyman, J. Libuda, *J. Mater. Chem. A* **2017**, *5*, 9250.
[5] Y. Long, S. Song, J. Li, L. Wu, Q. Wang, Y. Liu, R. Jin, H. Zhang, *ACS Catal.* **2018**, *8*, 8506.
[6] Q. Zhang, J. Bu, J. Wang, C. Sun, D. Zhao, G. Sheng, X. Xie, M. Sun, L. Yu, *ACS Catal.* **2020**, *10*, 10350.
[7] L. Pastor-Pérez, E. V. Ramos-Fernández, A. Sepúlveda-Escribano, *Int. J. Hydrogen Energy* **2019**, *44*, 21837.
[8] A. Mahata, A. S. Nair, B. Pathak, *Catal. Sci. Technol.* **2019**, *9*, 4835.
[9] M. Kourtelesis, T. S. Moraes, L. V. Mattos, D. K. Niakolas, F. B. Noronha, X. Verykios, *Appl. Catal., B* **2021**, *284*, 119757.
[10] T. Tong, Q. Xia, X. Liu, Y. Wang, *Catal. Commun.* **2017**, *101*, 129.
[11] H. Eliasson, Y. Niu, R. E. Palmer, H. Grönbeck, R. Erni, *Nanoscale* **2023**, *15*, 19091.
[12] Y. Gao, W. Wang, S. Chang, W. Huang, *ChemCatChem* **2013**, *5*, 3610.
[13] S. Xie, L. Liu, Y. Lu, C. Wang, S. Cao, W. Diao, J. Deng, W. Tan, L. Ma, S. N. Ehrlich, Y. Li, Y. Zhang, K. Ye, H. Xin, M. Flytzani-Stephanopoulos, F. Liu, *J. Am. Chem. Soc.* **2022**, *144*, 21255.
[14] G. Ferré, M. Aouine, F. Bosselet, L. Burel, F. J. Cadete Santos Aires, C. Geantet, S. Ntais, F. Maurer, M. Casapu, J. D. Grunwaldt, T. Epicier, S. Loridant, P. Vernoux, *Catal. Sci. Technol.* **2020**, *10*, 3904.
[15] P. D. Nellist, S. J. Pennycook, *Science* **1996**, *274*, 413.
[16] S. J. Pennycook, L. A. Boatner, *Nature* **1988**, *336*, 565.
[17] O. L. Krivanek, M. F. Chisholm, V. Nicolosi, T. J. Pennycook, G. J. Corbin, N. Dellby, M. F. Murfitt, C. S. Own, Z. S. Szilagyi, M. P. Oxley, S. T. Pantelides, S. J. Pennycook, *Nature* **2010**, *464*, 571.
[18] A. De Backer, S. Bals, S. Van Aert, *Ultramicroscopy* **2023**, *247*, 113702.
[19] A. De Backer, K. H. W. van den Bos, W. van den Broek, J. Sijbers, S. Van Aert, *Ultramicroscopy* **2016**, *171*, 104.
[20] J. M. LeBeau, S. D. Findlay, L. J. Allen, S. Stemmer, *Nano Lett.* **2010**, *10*, 4405.
[21] J. Madsen, P. Liu, J. Kling, J. B. Wagner, T. W. Hansen, O. Winther, J. Schiøtz, *Adv. Theory Simul.* **2018**, *1*, 1800037.
[22] J. M. LeBeau, S. D. Findlay, L. J. Allen, S. Stemmer, *Phys. Rev. Lett.* **2008**, *100*, 206101.
[23] J. M. Cowley, A. F. Moodie, *Acta Crystallogr.* **1957**, *10*, 609.
[24] J. Barthel, *Ultramicroscopy* **2018**, *193*, 1.
[25] O. Ronneberger, presented at *18th Int. Conf. on Medical Image Comput. and Computer Assisted Interventions (MICCAI)*, Munich, October, **2015**.
[26] J. Y. Zhu, presented at *2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, October, **2017**.
[27] F. Wang, T. R. Henninen, D. Keller, R. Erni, *Appl. Microsc.* **2020**, *50*, 23.
[28] Y. Liu, A. Chen, H. Shi, S. Huang, W. Zheng, Z. Liu, Q. Zhang, X. Yang, *Comput. Med. Imaging Graph.* **2021**, *91*, 101953.
[29] X. Li, G. Zhang, H. Qiao, F. Bao, Y. Deng, J. Wu, Y. He, J. Yun, X. Lin, H. Xie, H. Wang, Q. Dai, *Light. Sci. Appl.* **2021**, *10*, 44.
[30] A. Khan, C.-H. Lee, P. Y. Huang, B. K. Clark, *Npj Comput. Mater.* **2023**, *9*, 85.
[31] T. R. Henninen, M. Bon, F. Wang, D. Passerone, R. Erni, *Angew. Chem., Int. Ed.* **2020**, *59*, 839.
[32] E. Arslan Irmak, P. Liu, S. Bals, S. Van Aert, *Small Methods* **2021**, *5*, 2101150.
[33] R. F. Loane, P. Xu, J. Silcox, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 267.
[34] C. Ophus, *Adv. Struct. Chem. Image* **2017**, *3*, 13.
[35] L. Rangel DaCosta, H. G. Brown, P. M. Pelz, A. Rakowski, N. Barber, P. O'Donovan, P. McBean, L. Jones, J. Ciston, M. C. Scott, C. Ophus, *Micron* **2021**, *151*, 103141.
[36] I. Gulrajani, presented at *The 31st Int. Conf. on Neural Information Processing Systems (NIPS)*, Long Beach, (accessed: December 2017).
[37] T. Miyato, presented at *The 6th Int. Conf. on Learning Representations (ICLR)*, Vancouver, April, **2018**.
[38] A. Shrivastava, presented at *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Venice, (accessed: July 2017).
[39] D. Ulyanov, presented at *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, (accessed: July 2017).
[40] P. Isola, presented at *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, (accessed: July 2017).
[41] K. He, presented at *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, June, **2016**.
[42] R. Zhang, presented at *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, June, **2018**.
[43] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, (Preprint) arXiv:1602.07360, v4, submitted: Nov **2016**.
[44] M. Heusel, presented at *The 31st Annual Conf. on Neural Information Processing Systems (NIPS)*, Long Beach, December, **2017**.
[45] H. Eliasson, R. Erni, *npj Comput. Mat.* **2024**, *10*, 168.
[46] J. Uhrig, presented at *2017 Int. Conf. on 3D Vision (3DV)*, Qingdao, October, **2017**.
[47] A. Dosovitskiy, presented at *The 9th Int. Conf. on Learning Representations (ICLR), Virtual*, May, **2021**.
[48] A. Bruix, J. A. Rodriguez, P. J. Ramírez, S. D. Senanayake, J. Evans, J. B. Park, D. Stacchiola, P. Liu, J. Hrbek, F. Illas, *J. Am. Chem. Soc.* **2012**, *134*, 8968.
[49] J. L. Vincent, P. A. Crozier, *Nat. Commun.* **2021**, *12*, 5789.
[50] J. Quinson, M. Inaba, S. Neumann, A. A. Swane, J. Bucher, S. B. Simonsen, L. Theil Kuhn, J. J. K. Kirkensgaard, K. M. Ø. Jensen, M. Oezaslan, S. Kunz, M. Arenz, *ACS Catal.* **2018**, *8*, 6627.

