Deep Denoising for Scientific Discovery: A Case Study in Electron Microscopy

Anonymous authors
Paper under double-blind review

Abstract

Denoising is a fundamental challenge in scientific imaging. Deep convolutional neural networks (CNNs) provide the current state of the art in denoising natural images, where they produce impressive results. However, their potential has barely been explored in the context of scientific imaging. Denoising CNNs are typically trained on real natural images artificially corrupted with simulated noise. In contrast, in scientific applications, noiseless ground-truth images are usually not available. To address this issue, we propose a simulation-based denoising (SBD) framework, in which CNNs are trained on simulated images. We test the framework on data obtained from transmission electron microscopy (TEM), an imaging technique with widespread applications in material science, biology, and medicine. SBD outperforms existing techniques by a wide margin on a simulated benchmark dataset, as well as on real data. Apart from the denoised images, SBD generates likelihood maps to visualize the agreement between the structure of the denoised image and the observed data. Our results reveal shortcomings of state-of-the-art denoising architectures, such as their small field-of-view: substantially increasing the field-of-view of the CNNs allows them to exploit non-local periodic patterns in the data, which is crucial at high noise levels. In addition, we analyze the generalization capability of SBD, demonstrating that the trained networks are robust to variations of imaging parameters and of the underlying signal structure. Finally, we release the first publicly available benchmark dataset of TEM images, containing 18,000 examples.

1 Introduction

Imaging technology is an essential tool in most scientific domains. Electron microscopy enables the visualization of atomic structures (Zuo & Spence, 2017), fluorescence microscopy makes it possible to study cellular processes (Lichtman & Conchello, 2005), telescopes reveal galaxies and other astronomical objects that are light years away (McLean, 2008). In all these imaging modalities, images are corrupted by noise associated with stochastic processes occurring during signal generation and detection, degrading the information content of the data. The goal of denoising is to estimate and restore the missing information from the image.

In the past decade, convolutional neural networks (CNNs) (LeCun et al., 2015) have achieved state-of-the-art performance in image denoising (Zhang et al., 2017; Chen & Pock, 2016). However, the potential of this methodology has barely been explored in the context of scientific imaging. The vast majority of published work concentrates on visual images, training the CNNs using photographs corrupted with simulated additive Gaussian noise. In contrast, noiseless ground-truth images are usually not available in scientific applications. To address this issue, we propose a simulation-based denoising (SBD) framework, in which CNNs are trained on simulated images. We validate our methodology through a case study in transmission electron microscopy.

Transmission electron microscopy (TEM) is a powerful and versatile characterization technique used to probe the atomic-level structure and composition of a wide range of materials, such as catalysts or semiconductors (Smith, 2015; Tao & Crozier, 2016). It has had a huge impact in structural biology as recognized with the award of the 2017 Nobel Prize in Chemistry (Bai et al., 2015). Recent advancements in direct electron detection systems enable experimentalists to image dynamic events at frame rates in the kilohertz range (Faruqi & McMullan, 2018; Ercius et al., 2020). Imaging at these
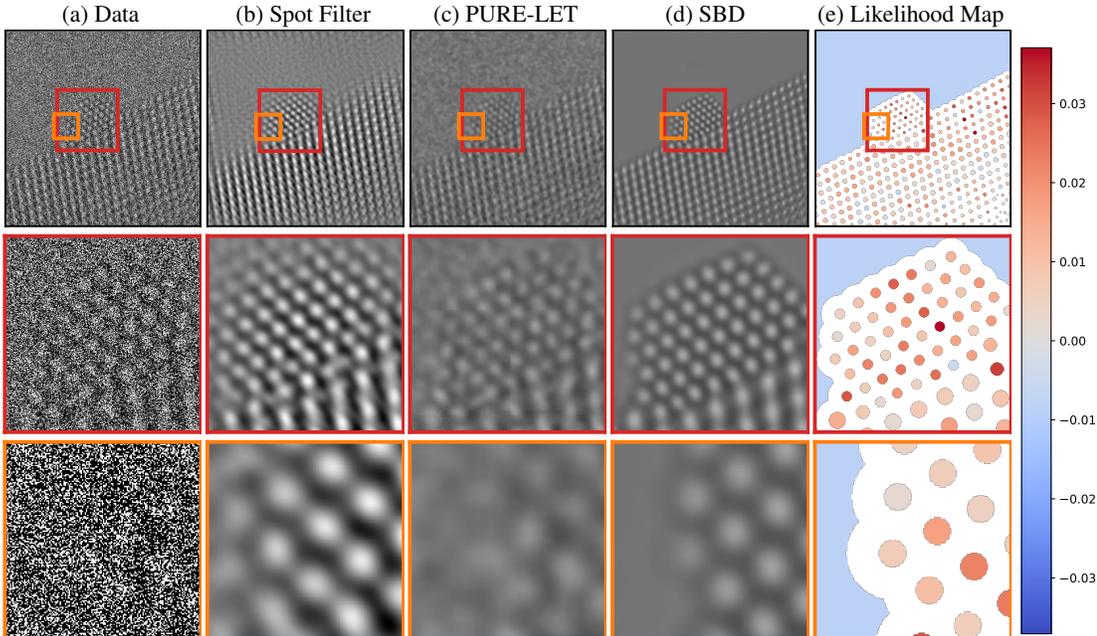
Figure 1: **Denoising results for real data.** (a) An atomic-resolution electron-microscope image of a platinum nanoparticle obtained via transmission electron microscopy at a magnification of over one million, as described in Section 4. The average image intensity is 0.45 electrons/pixel (i.e. a large fraction of pixels represent zero electrons!), which results in an extremely low signal-to-noise ratio. (b) Denoised image obtained via Fourier-based filtering by a domain expert. (c) Denoised image obtained via the wavelet-based PURE-LET method (Luisier et al., 2010). (d) Denoised image obtained by the proposed Simulation-Based Denoising framework. (e) Likelihood map quantifying to what extent the atomic structure identified from the SBL denoised image is consistent with the data (see Section 3). Regions in red are more likely to correspond to atomic columns in the nanoparticle. Regions in blue are more likely to belong to the vacuum.

time scales is critical to advance our understanding of functional materials. In catalytic systems, for example, the chemical transformation process is accompanied by dynamic, atomic-level structural rearrangements which may occur over a time scale spanning tens of milliseconds (Sun et al., 2020; Guo et al., 2020; Lawrence et al., 2020; Levin et al., 2020). Unfortunately, acquiring image series at such high temporal resolution necessarily produces datasets that are severely degraded by shot noise, rendering traditional imaging processing approaches ineffective. It is typically not feasible to reduce the noise content of the frames in the movie by increasing the intensity of the incident electron beam, since the high-energy beam used can seriously damage the material at high doses. Consequently, there is an acute need for novel denoising technology in this domain.

The main contributions of this work are the following. First, we introduce a general framework to perform deep-learning based denoising on scientific-imaging data. The framework is applicable beyond electron microscopy, in any domain where the objects of interest can be simulated, such as medical imaging (Kim et al., 2019; Minarik et al., 2020), other types of microscopy (Giannatou et al., 2019; Vasudevan & Jesse, 2019), or astronomy (Peterson et al., 2015). The proposed denoising method outperforms existing techniques by a wide margin on a benchmark dataset of TEM images, as well as on real TEM data (see Figure 1 and Section 6). Second, we propose a method to visualize the output of the image in the form of likelihood maps that quantify the agreement between the estimated image and the observed data. Third, we perform a thorough analysis of the generalization capability of our method, demonstrating that the trained networks are robust to variations of imaging parameters and of the underlying signal structure. Our results demonstrate that architectures optimized for natural photographic images may have fundamental shortcomings when applied to domain-specific data. For instance, a gradient-based analysis shows that substantially increasing the field-of-view of the networks allows them to exploit non-local periodic patterns in the data, resulting in a significant
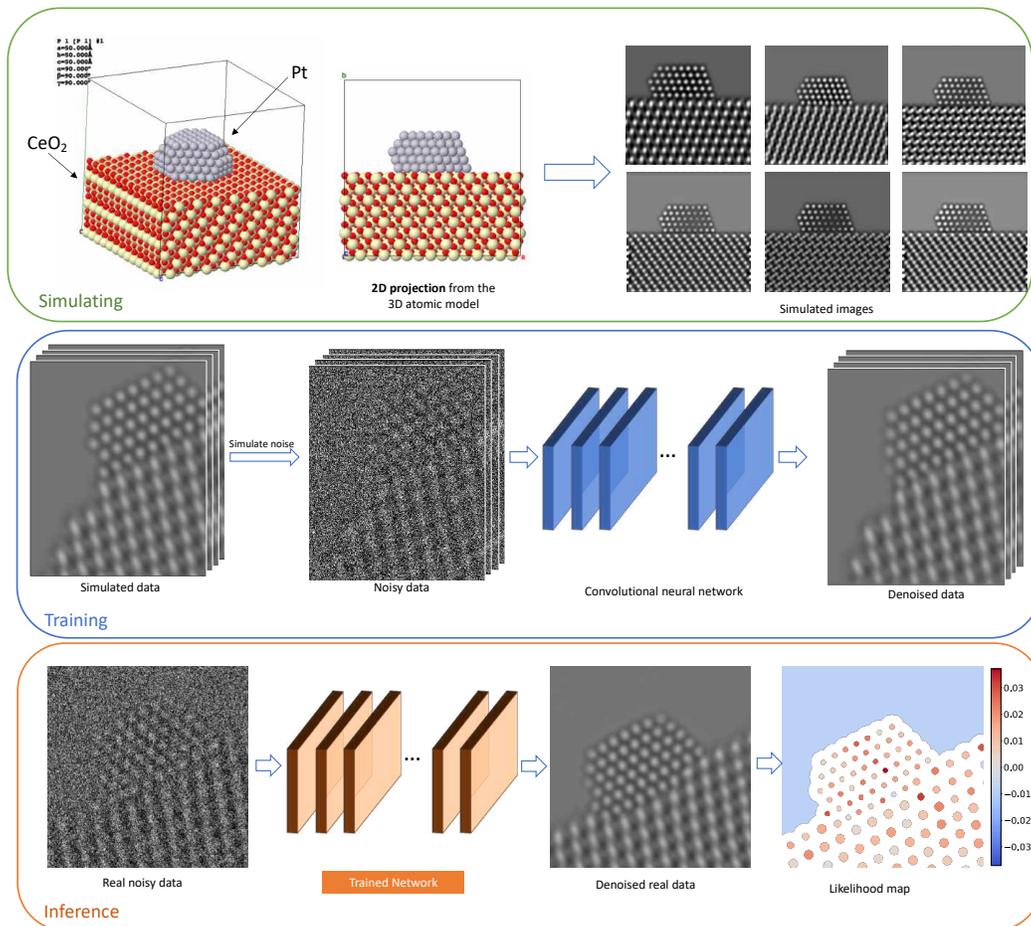
Figure 2: **Simulation-based denoising framework**. (Top) A training dataset is generated by simulating images with different structures at varying imaging conditions. (Middle) A CNN is trained using the simulated images, paired with noisy counterparts obtained by simulating the relevant noise process. (Bottom) The trained CNN is applied to real data to yield a denoised image. After analyzing the image to extract structure of interest, a likelihood map is generated to quantify the agreement between this structure and the noisy data.

boost in performance. Finally, to encourage further development of deep-learning methodologies for scientific imaging, we release the first publicly available denoising benchmark dataset of TEM images, containing 18,000 examples.

2 RELATED WORK

A wide variety of denoising methods have been applied across different scientific imaging modalities, including traditional linear filters (Nellist & Pennycook, 1998), nonlinear filters (Tomasi & Manduchi, 1998; Milanfar, 2012; Jiang et al., 2003), wavelet-based methods (Chang et al., 2000; Portilla et al., 2003; Zhu et al., 2015; Meiniel et al., 2018), and sparsity-based approaches (Meiniel et al., 2018; Beckouche et al., 2013). Deep convolutional networks have been shown to outperform all of these approaches (Zhang et al., 2017; Chen & Pock, 2016), but the rapidly growing literature on this methodology focuses almost exclusively on photographic images. We are aware of only a few very recent exceptions. In the medical domain, Kim et al. (2019), Gong et al. (2018), and Minarik et al. (2020) apply CNN-based denoising to low-dose computer tomography, positron-emission tomography and scintillation-camera data respectively. Giannatou et al. (2019), Vasudevan & Jesse (2019), and Ede & Beanland (2019) apply CNNs to denoise simulated scanning electron microscopy
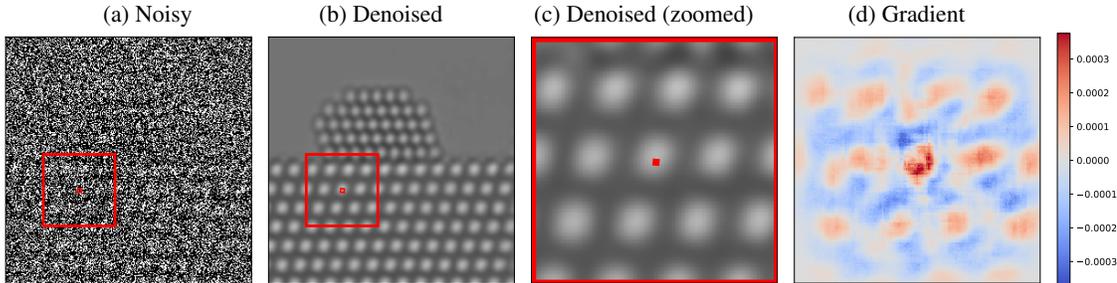
| (a) Noisy | (b) Denoised | (c) Denoised (zoomed) | (d) Gradient |

Figure 3: **Gradient analysis of the learned denoising function.** (a) To compute the red pixel in the denoised image (b), the proposed CNN uses a 220×220 area (red box) in the noisy image (a). The gradient of the denoised pixel with respect to its input indicates what regions in the noisy image have a greater influence on the estimate (according to a first-order Taylor approximation to the denoising map). The gradient (d) weights nearby pixels more heavily, but also has significant magnitude at pixels located on different atoms. This suggests that the CNN combines local and non-local information to estimate the pixel. See Figure 18 for additional examples using real data.

data, without validating on real data. Manifold et al. (2019) train CNNs to denoise Raman scattering data, using measurements gathered at a higher signal-to-noise ratio (SNR) as ground-truth images. These results showcase the potential of deep denoising for scientific imaging, but also the challenges of gathering adequate datasets to train the deep networks. In this work, we show that networks trained on simulated data can generalize effectively to real data, which may enable the application of deep denoising in domains where it is infeasible to measure data at high SNRs.

Unsupervised denoising is a promising alternative for applications where no ground-truth images are available. Unsupervised wavelet methods achieve performance comparable to their supervised counterparts (Luisier et al., 2007; Hel-Or & Shaked, 2008; Raphan & Simoncelli, 2008). Noise2Noise (Lehtinen et al., 2018), a deep-learning approach that requires access to pairs of noisy images corresponding to the same underlying signal, has been applied to cryo electron microscopy by Buchholz et al. (2019). More recent methods (Krull et al., 2019a; Batson & Royer, 2019; Laine et al., 2019) can be trained directly on noisy images. Krull et al. (2019b) Khademi et al. (2020), Prakash et al. (2020) and Zhang et al. (2019) apply this approach to fluorescence microscopy data. In the case of the TEM data in our case study, unsupervised denoising does not perform well (see Section 6, and in particular Figures 16, and 17), possibly due to the SNR, which is orders of magnitude lower (around 3 dB) than that reported in these works (around 27 dB for Zhang et al. (2019)). Improving the performance of unsupervised denoising methods at low SNRs is thus an important topic for future research - Wang et al. (2020) have reported promising results in this direction for scanning TEM.

3 METHODOLOGY

SIMULATION-BASED DENOISING

Current state-of-the-art deep-learning techniques for denoising photographic images require a training set of ground-truth images (Zhang et al., 2017). The CNNs are trained by corrupting these clean images, typically with additive Gaussian noise, and then minimizing the mean squared error between the network output and the original images. The main obstacle to leveraging this approach in scientific imaging is the lack of ground-truth data; in many applications there is no such thing as a *clean image*. We address this by using a dataset of simulated images to train the CNNs. We call this framework simulation-based denoising (SBD). The diagram in Figure 2 shows the different stages: simulation of the training set, training of the CNNs using the simulated data, and inference on the real data.

A crucial difference between SBD and previous methodology for deep denoising is that the procedure for generating the training set needs to be explicitly *designed*. In order to ensure effective generalization to real data, we must include sufficient variation of imaging parameters and image structure in the training dataset. In addition, particular care is needed to enforce invariance to small changes in the geometry of the image. Figure 12 shows that a denoising CNN can easily overfit the specific
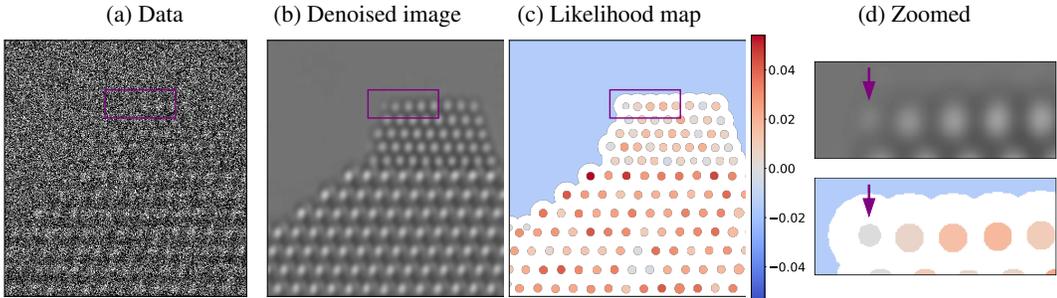
Figure 4: **Likelihood map.** When the simulated noisy image in (a) is denoised using the proposed framework (b), a spurious atom appears at the left edge of the nanoparticle (see zoomed image (d)). The likelihood map (c) at that location is negative, which indicates that the presence of an atom is less likely than its absence according to the observed data.

alignment and scale of the training data. This issue can be addressed by augmenting the training set with rotated and scaled versions of the simulated images.

We provide a detailed account of the simulated dataset for our case study in Sections 4 and A.1 and of the noise model in Section B. The resulting CNNs generalize well to the real data, as well as across different imaging parameters and signal structures (see Section 6). Determining how to optimally sample the space of possible simulation parameters when generating data to train CNNs for denoising is an important methodological question for future research.

EXPLOITING NON-LOCAL SIGNAL STRUCTURE

Images in scientific applications often have a distribution of pixel intensities that is very different to that of natural images. Our case study shows that it is crucial to take this into account in order to achieve successful denoising. Current state-of-the-art networks for denoising natural images have small fields of view. For example, the field of view of networks used in DnCNN (Zhang et al., 2017) is 41×41 pixels, and that of DURR (Zhang et al., 2018) is 45×45 pixels. This is sufficient in the case of most natural images, because they have rich local structure and the SNR of interest is relatively high (most works focus on an SNR above 22 dB, see e.g. Zhang et al. (2017)). In contrast, microscopy images often contain periodic structure, and are measured at extremely low SNRs (in our case, the SNR for the real TEM data is about 3 dB).

Mohan et al. (2020) have shown that denoising CNNs tend to average over larger regions of surrounding pixels as the SNR decreases (qualitatively, this is the same behavior observed in a classical linear Wiener filter). Motivated by this, we propose a modified UNet architecture, described in Section C, with a field of view of 220×220 pixels (i.e. a 25-fold increase in area with respect to generic denoising architectures). This results in a very significant boost in performance (more than 7dB in PSNR, see Table 1). Table 1 shows a more detailed comparison to other CNN architectures, which suggests that the larger field of view is the reason for the improvement. Increasing the field of view of the UNet produces a dramatic increase of around 6 dB in PSNR, even if the number of parameters remains similar. Following Mohan et al. (2020), we perform a gradient-based analysis to visualize the linear term in the first-order Taylor decomposition of the denoising map with respect to its input for specific pixels. The analysis reveals that the denoising strategy learned by the proposed network simultaneously exploits local and non-local structure in the noisy data (see Figures 3 and 18). This showcases the remarkable flexibility of data-driven denoising based on deep learning.

LIKELIHOOD MAPS

In most applied domains, the goal of denoising is to uncover image structure of scientific interest. In our case study, this corresponds to the location and intensity of projected columns of atoms in a catalytic nanoparticle that is surrounded by a vacuum. Quantifying to what extent such structure is consistent with the observed measurements is therefore of great interest. We propose to achieve this

by computing the likelihood of the data *with respect to meaningful features identified in the denoised image*. The general procedure, and its implementation in the case of our case study, are as follows:

1. Identify a region of interest $\mathcal{R}$. In our case study, this would correspond to an atomic column, located for example via blob detection, or to the vacuum.

2. Fit a low-dimensional model to the denoised image within the region of interest, which yields an estimate x_i associated to each pixel location $i \in \mathcal{R}$. In our case, we assume that the intensity of the atomic columns and the vacuum are constant, so the estimate is obtained by averaging over all denoised pixels in $\mathcal{R}$.

3. Compute the likelihood of the noisy data in $\mathcal{R}$ with respect to the estimated pixel values. In our case, the noise is approximately iid Poisson (see Section B), so the likelihood is given by

$$\mathcal{L}(R) := \prod_{i \in R} p_{x_i}(y_i), \tag{1}$$

where y_i denotes the noisy value in the ith pixel, and p_{x_i} is a Poisson pmf with rate parameter x_i.

This technique makes it possible to consider different hypotheses about the underlying image structure and compare their agreement with the observed data. In our case study, we evaluate the hypotheses that a detected atomic column is (1) truly there, or (2) an artifact introduced by the denoising procedure. The likelihood under hypothesis (1) is computed as above. The likelihood under hypothesis (2) is computed by setting the estimate x_i to equal to the average intensity of the noisy pixels identified as belonging to the vacuum region. To visualize the consistency of the two hypotheses with the measured data, we plot the difference in their log likelihood for each region of interest. This is equivalent to performing a log-likelihood ratio test. We call this visualization a *likelihood map*. Figures 1 and 4 show likelihood maps for the real data and for a simulated example. In the simulated example, a spurious atom is detected at the left end of the zoomed region. However, the likelihood map at that location is negative, which correctly indicates that the presence of an atom is less likely than its absence according to the observed data.

4 Dataset

The TEM image data used in this work correspond to images from a widely utilized catalytic system, which consists of platinum (Pt) nanoparticles supported on a larger cerium (IV) oxide (CeO$_2$) nanoparticles. This bi-functional catalytic system is ubiquitously used in clean energy conversion and environmental remediation applications, in addition to a broad range of other chemical reactions (Montini et al., 2016; Yu et al., 2012; Nie et al., 2015). From a general point of view, this system can be considered as a model for supported nanoparticle catalysts, since a large number of inorganic catalysts are based on small metallic nanoparticles supported over different oxides. Thus, results and conclusions extracted from the current work are relevant to a great number of similar samples in the field of catalysis (e.g., oxide crystals supporting metal nanoparticles). The real data used to test the proposed SBD framework consists of a series of images of the Pt/CeO$_2$ catalyst. The images were acquired in a N$_2$ gas atmosphere using an aberration-corrected FEI Titan transmission electron microscope (TEM), operated at 300 kV and coupled with a Gatan K2 IS direct electron detector. The detector was operated in electron counting mode with a time resolution of 0.025 sec/frame and an incident electron dose rate of 5,000 e$^-$/Å^2/s. The electromagnetic lens system of the microscope was tuned to achieve a highly coherent parallel beam configuration with minimal low-order aberrations (e.g., astigmatism, coma), and a third-order spherical aberration coefficient of approximately -13 μm.

The simulated TEM image dataset was generated using the multi-slice image simulation method, as implemented in the Dr. Probe software package (Barthel, 2018) (see Section A.1 for more details on the simulation process). Images were simulated with 1024 x 1024 pixels and then binned to match the approximate pixel size of the experimentally acquired image series. To equate the intensity range of the simulated images with those acquired experimentally, the intensities of the simulated images were scaled by a factor which equalized the vacuum intensity in a single simulation to the average intensity measured over a large area of the vacuum in a single 0.025 second experimental frame (i.e., 0.45 counts per pixel in the vacuum region).

In the type of TEM imaging performed in this work, multiple electron-optical and specimen parameters can give rise to complex, non-linear modulations of the image contrast. These parameters

METHODS	PSNR	SSIM
Raw	3.56 ± 0.03	0.00 ± 0.00
Low Pass Filter (Nellist & Pennycook, 1998)	21.59 ± 0.07	0.44 ± 0.03
Adaptive Wiener Filter (Lim, 1990)	22.42 ± 1.08	0.63 ± 0.02
VST + NLM (Buades et al., 2005)	26.55 ± 0.16	0.73 ± 0.01
VST + BM3D (Makitalo & Foi, 2012)	25.27 ± 0.15	0.80 ± 0.01
PURE-LET (Luisier et al., 2010)	28.36 ± 0.88	0.93 ± 0.01
SBD + DnCNN (Zhang et al., 2017)	30.47 ± 0.64	0.93 ± 0.01
SBD + Small UNet (Zhang et al., 2018)	30.87 ± 0.56	0.93 ± 0.01
SBD + Proposed Architecture	$\mathbf{38.05} \pm 0.81$	$\mathbf{0.99} \pm 0.01$

Table 1: **Results on simulated test data.** Mean PSNR and SSIM ($\pm$ standard deviation) of different denoising methods on the held-out simulated test set described in Section 5. SBD approaches achieve the best results. SBD combined with the proposed architecture outperforms all other techniques by more than 7 dB.

include the objective lens defocus, the specimen thickness, the orientation of the specimen, and its crystallographic shape/structure. Various combinations of these parameters may cause the contrast of atomic columns in the image to appear as black, white, or an intermediate mixture of the two (see, e.g., Figure 6). When designing the simulated dataset for the SBD framework, it is necessary to include images simulated under widely varied conditions, in order to cover the breadth of possibilities which may arise during a typical experiment. A skilled microscopist attempts to acquire images under conditions in which the image contrast can be interpreted, which limits the overall size of the parameter space under consideration. However, various instances of defocus, tilt, thickness, and shape/structure inevitably arise. To generate our dataset we systematically varied these parameters to produce a large number of potential combinations (approximately 18,000), as described in Sections A.2 and A.3.

5 EXPERIMENTS

As explained in Section 3 our proposed CNN is a modified UNet with a very large field of view (see Section C.1 for a more detailed description of the architecture and training procedure). In order to study the generalization ability of the CNN across different imaging parameters and signal structures we divided the simulated dataset described in Section 4 into different subsets. These subsets were classified based on (1) the character of the atomic column contrast, (2) the structure/size of the supported Pt nanoparticle, and (3) the defects of the Pt surface structure. The contrast was classified into three divisions, black, intermediate, or white contrast, by a domain expert (see Figure 6). The nanoparticle structure was classified into four categories, "PtNp1" through "PtNp4". PtNp1 and PtNp2 correspond to supported Pt nanoparticles 2 nm in size where the difference is the appearance of an atomic column located at the interface between the Pt and the CeO_2 support; PtNp3 corresponds to a Pt nanoparticle 1 nm in size; and PtNp4 corresponds to a Pt nanoparticle 3 nm in size. Finally, the defects were divided into five categories: "D0", "D1", "D2", "Dh", and "Ds" in accordance with the atomic-scale structural models presented in A.1 and in particular in Figure 10. D0 is the initial structure, D1/D2 a structure in which 1/2 atoms have been removed respectively, Dh a structure in which a column has been reduced to half its original occupancy, and Ds a structure in which a column has been reduced to a single atom. The generalization ability of the proposed CNN was evaluated by systematically training on each of the subsets and testing on the rest. The number of images in each subset was fixed to be equal in order to ensure a fair comparison.

The imaging parameters of the real-data collection, described in Section 4, correspond to the white contrast category. We therefore used the subset of simulated dataset corresponding to this contrast (5583 images) to train the proposed CNN. 90% of the data were used for training. The remaining 559 images were evenly split into validation and test sets. We also trained two state-of-the-art architectures for natural image denoising (Zhang et al., 2017; 2018) (see Sections C.2 and C.3). We compared the SBD approaches to a variety of popular denoising methods: low-pass filtering (Nellist & Pennycook, 1998), adaptive Wiener filtering (Lim, 1990), BM3D (Makitalo & Foi, 2012), non-local means (Luisier et al., 2010), a wavelet-based method known as PURE-LET (Luisier et al., 2010), and an unsupervised deep learning approach based on a blind-spot CNN (Laine et al., 2019). A detailed description of these techniques is provided in Section D. All hyperparameters were chosen based
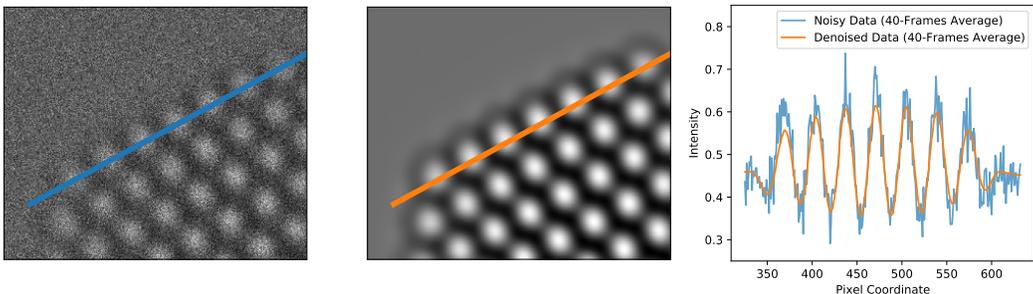
Figure 5: **Validation on real data**. The real data consist of 40 frames which are approximately stationary and aligned. Their temporal estimate therefore provide a reasonable estimate for the true intensity profile. We compare the average intensity profile on the surface atomic columns of the platinum for the denoised and noisy data. The profiles are very similar (except for some spurious fluctuations in the average of the noisy data), which suggests that the proposed approach achieves effective denoising on the real data.

on the validation data. Performance was measured in terms of SSIM (Wang et al., 2004) and peak signal-to-noise ratio (PSNR). The methods were also applied to the real data.

6 RESULTS AND DISCUSSION

The results of our computational experiments demonstrate that simulation-based denoising is an effective denoising methodology for TEM data. Our proposed CNN outperforms all other methods by a margin of 7 dB in PSNR on the simulated test data, as shown in Table 1, and Figures 14 and 15. SBD recovers the overall shape of the nanoparticle, the interface between the nanoparticle and the support, and the different periodic patterns of the CeO2 support and Pt nanoparticle. The image contrast from the CeO2 shows a subtle pattern of bright, intermediate and dark features which are associated with different components of the atomic structure of the crystal. These contrast features are well reproduced in the images denoised via SBD but are mostly absent from the other baseline approaches. Results on real data are shown in Figures 1, 16, and 17. SBD produces denoised images that are of much higher quality than those of other methods, which contain obvious artifacts. Figure 5 further validates the denoising results of our proposed method on the real data by using a temporal average to approximate a high SNR image. The performance of SBD is robust to variations in imaging parameters and in the underlying signal structure, as shown in Figure 13. We only observe a significant decrease in performance when the network is trained on black-contrast images (defined in Section 5) and tested on other contrasts (interestingly the network generalizes well from white and intermediate contrasts to black contrasts).

Our case study is a proof of concept that CNNs trained on simulated data can be remarkably effective in improving results on real data. It provides several generalizable insights for the application of this methodology in scientific imaging, and also suggests avenues for future research. First, the design of the training dataset is absolutely crucial: The framework may fail if it encounters features that are absent from the training set. An important question is how to design simulated datasets in a principled systematic way. Answering it will require a deeper understanding of the generalization ability of CNNs with respect to variations in the statistics of the input images. Second, we demonstrated that architectures tailored to photographic imaging can perform poorly when applied to other data. Designing CNNs for other domains requires understanding what image features are being exploited. Gradient visualization is shown to be useful here, but more advanced visualization techniques are needed. Third, although SBD outperform other methods by a large margin, some artifacts such as phantom atoms still appear. Our proposed likelihood maps help to flag such events, but may still fail to do so in regions of unusually low SNR. Developing more sophisticated methods for uncertainty quantification is therefore a key research direction. It would also be of great interest to develop unsupervised or self-supervised denoising approaches that are effective with small amounts of data at low SNRs. Finally, to encourage further development of deep-learning methodologies for scientific imaging, we release a denoising benchmark dataset of TEM images, containing 18,000 examples.

REFERENCES

Xiao-Chen Bai, Greg McMullan, and Sjors H.W Scheres. How cryo-em is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1):49 – 57, 2015. ISSN 0968-0004. doi: https://doi.org/10.1016/j.tibs.2014.10.005. URL `http://www.sciencedirect.com/science/article/pii/S096800041400187X`.

Juri Barthel. Dr. probe: A software for high-resolution stem image simulation. *Ultramicroscopy*, 193: 1–11, 2018.

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. *arXiv preprint arXiv:1901.11365*, 2019.

Simon Beckouche, Jean-Luc Starck, and Jalal Fadili. Astronomical image denoising using dictionary learning. *Astronomy & Astrophysics*, 556:A132, 2013.

S Bernal, FJ Botana, JJ Calvino, C Lopez-Cartes, JA Perez-Omil, and JM Rodrıguez-Izquierdo. The interpretation of hrem images of supported metal catalysts using image simulation: profile view images. *Ultramicroscopy*, 72(3-4):135–164, 1998.

Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pp. 60–65. IEEE, 2005.

Tim-Oliver Buchholz, Mareike Jordan, Gaia Pigino, and Florian Jug. Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 502–506. IEEE, 2019.

S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Processing*, 9(9):1532–1546, 2000.

Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2016.

Jeffrey M Ede and Richard Beanland. Improving electron micrograph signal-to-noise with an atrous convolutional encoder-decoder. *Ultramicroscopy*, 202:18–25, 2019.

Peter Ercius, Ian Johnson, Hamish Brown, Philipp Pelz, Shang-Lin Hsu, Brent Draney, Erin Fong, Azriel Goldschmidt, John Joseph, Jason Lee, and et al. The 4d camera – a 87 khz frame-rate detector for counted 4d-stem experiments. *Microscopy and Microanalysis*, pp. 1–3, 2020. doi: 10.1017/S1431927620019753.

A.R. Faruqi and G. McMullan. Direct imaging detectors for electron microscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 878:180 – 190, 2018. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2017.07.037. URL `http://www.sciencedirect.com/science/article/pii/S0168900217307787`. Radiation Imaging Techniques and Applications.

E Giannatou, G Papavieros, V Constantoudis, H Papageorgiou, and E Gogolides. Deep learning denoising of sem images towards noise-reduced ler measurements. *Microelectronic Engineering*, 216:111051, 2019.

Kuang Gong, Jiahui Guan, Chih-Chieh Liu, and Jinyi Qi. Pet image denoising using a deep neural network through fine tuning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2): 153–161, 2018.

Han Guo, Philippe Sautet, and Anastassia N. Alexandrova. Reagent-triggered isomerization of fluxional cluster catalyst via dynamic coupling. *The Journal of Physical Chemistry Letters*, 11 (8):3089–3094, 2020. doi: 10.1021/acs.jpclett.0c00548. URL `https://doi.org/10.1021/acs.jpclett.0c00548`. PMID: 32227852.

Y. Hel-Or and D. Shaked. A discriminative approach for wavelet denoising. *IEEE Transactions on Image Processing*, 17:443–457, 2008.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

Wen Jiang, Matthew L Baker, Qiu Wu, Chandrajit Bajaj, and Wah Chiu. Applications of a bilateral denoising filter in biological electron microscopy. *Journal of structural biology*, 144(1-2):114–122, 2003.

Wesley Khademi, Sonia Rao, Clare Minnerath, Guy Hagen, and Jonathan Ventura. Self-supervised poisson-gaussian denoising. *arXiv preprint arXiv:2002.09558*, 2020.

Byeongjoon Kim, Minah Han, Hyunjung Shim, and Jongduk Baek. A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images. *Medical physics*, 46(9):3906–3923, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Earl J Kirkland et al. Image simulation in transmission electron microscopy. *Cornell University, Ithaca*, 2006.

Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137, 2019a.

Alexander Krull, Tomas Vicar, and Florian Jug. Probabilistic noise2void: Unsupervised content-aware denoising. *arXiv preprint arXiv:1906.00651*, 2019b.

Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. In *Advances in Neural Information Processing Systems*, pp. 6970–6980, 2019.

Ethan L. Lawrence, Barnaby D.A. Levin, Benjamin K. Miller, and Peter A. Crozier. Approaches to exploring spatio-temporal surface dynamics in nanoparticles with in situ transmission electron microscopy. *Microscopy and Microanalysis*, 26(1):86–94, 2020. doi: 10.1017/S1431927619015228.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

Barnaby D.A. Levin, Ethan L. Lawrence, and Peter A. Crozier. Tracking the picoscale spatial motion of atomic columns during dynamic structural change. *Ultramicroscopy*, 213:112978, 2020. ISSN 0304-3991. doi: https://doi.org/10.1016/j.ultramic.2020.112978. URL `http://www.sciencedirect.com/science/article/pii/S0304399119303122`.

Jeff W Lichtman and José-Angel Conchello. Fluorescence microscopy. *Nature methods*, 2(12):910–919, 2005.

Jae S Lim. Two-dimensional signal and image processing. *ph*, 1990.

F. Luisier, T. Blu, and M. Unser. A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, 16:593–606, 2007.

Florian Luisier, Thierry Blu, and Michael Unser. Image denoising in mixed poisson–gaussian noise. *IEEE Transactions on image processing*, 20(3):696–708, 2010.

Markku Makitalo and Alessandro Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE transactions on image processing*, 22(1):91–103, 2012.

Bryce Manifold, Elena Thomas, Andrew T Francis, Andrew H Hill, and Dan Fu. Denoising of stimulated raman scattering microscopy images via deep learning. *Biomedical optics express*, 10 (8):3860–3874, 2019.

Ian S McLean. *Electronic imaging in astronomy: detectors and instrumentation*. Springer Science & Business Media, 2008.

William Meiniel, Jean-Christophe Olivo-Marin, and Elsa D Angelini. Denoising of microscopy images: a review of the state-of-the-art, and a new sparsity-based method. *IEEE Transactions on Image Processing*, 27(8):3842–3856, 2018.

Peyman Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE signal processing magazine*, 30(1):106–128, 2012.

David Minarik, Olof Enqvist, and Elin Trägårdh. Denoising of scintillation camera images using a deep convolutional neural network: a monte carlo simulation approach. *Journal of Nuclear Medicine*, 61(2):298–303, 2020.

Sreyas Mohan, Zahra Kadkhodaie, Eero P. Simoncelli, and Carlos Fernandez-Granda. Robust and interpretable blind image denoising via bias-free convolutional neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HJlSmC4FPS.

Tiziano Montini, Michele Melchionna, Matteo Monai, and Paolo Fornasiero. Fundamentals and catalytic applications of ceo2-based materials. *Chemical reviews*, 116(10):5987–6041, 2016.

PD Nellist and SJ Pennycook. Accurate structure determination from image reconstruction in adf stem. *Journal of Microscopy*, 190(1-2):159–170, 1998.

Yao Nie, Li Li, and Zidong Wei. Recent advancements in pt and pt-free catalysts for oxygen reduction reaction. *Chemical Society Reviews*, 44(8):2168–2201, 2015.

JR Peterson, JG Jernigan, SM Kahn, AP Rasmussen, E Peng, Z Ahmad, J Bankert, C Chang, C Claver, DK Gilmore, et al. Simulation of astronomical images from optical survey telescopes using a comprehensive photon monte carlo approach. *The Astrophysical Journal Supplement Series*, 218 (1):14, 2015.

Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Processing*, 12(11), 2003.

Mangal Prakash, Manan Lalit, Pavel Tomancak, Alexander Krul, and Florian Jug. Fully unsupervised probabilistic noise2void. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 154–158. IEEE, 2020.

Martin Raphan and Eero P Simoncelli. Optimal denoising in redundant representations. *IEEE Transactions on image processing*, 17(8):1342–1352, 2008.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

David Smith. *CHAPTER 1: Characterization of nanomaterials using transmission electron microscopy*, pp. 1–29. Number 37 in RSC Nanoscience and Nanotechnology. Royal Society of Chemistry, 37 edition, January 2015. doi: 10.1039/9781782621867-00001.

Geng Sun, Anastassia N. Alexandrova, and Philippe Sautet. Structural rearrangements of subnanometer cu oxide clusters govern catalytic oxidation. *ACS Catalysis*, 10(9):5309–5317, 2020. doi: 10.1021/acscatal.0c00824. URL https://doi.org/10.1021/acscatal.0c00824.

Franklin Tao and Peter Crozier. Atomic-scale observations of catalyst structures under reaction conditions and during catalysis. *Chemical Reviews*, 116(6):3487–3539, March 2016. ISSN 0009-2665. doi: 10.1021/cr5002657.

Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *ICCV*, volume 98, 1998.

Rama K Vasudevan and Stephen Jesse. Deep learning as a tool for image denoising and drift correction. *Microscopy and Microanalysis*, 25(S2):190–191, 2019.

Feng Wang, Trond R Henninen, Debora Keller, and Rolf Erni. Noise2atom: Unsupervised denoising for scanning transmission electron microscopy images. 2020.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Weiting Yu, Marc D Porosoff, and Jingguang G Chen. Review of pt-based bimetallic catalysis: from model surfaces to supported catalysts. *Chemical reviews*, 112(11):5780–5817, 2012.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, 2017.

Xiaoshuai Zhang, Yiping Lu, Jiaying Liu, and Bin Dong. Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration. *arXiv preprint arXiv:1805.07709*, 2018.

Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11710–11718, 2019.

Hai Jing Zhu, Bo Chong Han, and Bo Qiu. Survey of astronomical image processing methods. In *International Conference on Image and Graphics*, pp. 420–429. Springer, 2015.

Jian-Min Zuo and J.C.H. Spence. *Advanced Transmission Electron Microscopy, Imaging and Diffraction in Nanoscience*. 01 2017. ISBN ISBN 978-1-4939-6607-3.

A   Data simulation

A.1   Simulation process

The simulated TEM image dataset was generated using the multi-slice image simulation method, as implemented in the Dr. Probe software package (Barthel, 2018). In the multi-slice approach, the modeled specimen is sectioned into many thin slices (here they are 0.167 Angstroms thick), and quantum mechanical calculations are performed to simulate the incident electron wave function propagating through and interacting with each slice of the material (Kirkland et al., 2006). The resultant wave function exiting the last slice is then convolved with a point spread function that emulates the effect of imaging it in the electron microscope. All of the image simulations were performed using an accelerating voltage of 300 kV with a beam convergence angle of 0.2 mrad and a focal spread of 4 nm. The third-order spherical aberration coefficient (C_s) was set to be -13 μm. The fifth-order spherical aberration coefficient ($C5$) was set as 5 mm. All other aberrations (e.g., 2-fold and 3-fold astigmatism, coma, star aberration, etc.) were approximated to be negligible. The defocus ($C1$) was varied systematically between 0 nm and 20 nm, as discussed below. Image calculations were computed using a non-linear model including partial temporal coherence by explicit averaging and partial spatial coherence, which is treated by a quasi-coherent approach with a dampening envelope applied to the wave function. An isotropic vibration envelope of 50 pm was applied during the image calculation. Images were simulated with 1024 x 1024 pixels and then later binned to desired sizes to match the pixel size of the experimentally acquired image series. Finally, to equate the intensity range of the simulated images with those acquired experimentally, the intensities of the simulated images were scaled by a factor which equalized the vacuum intensity in a single simulation to the average intensity measured over a large area of the vacuum in a single 0.025 second experimental frame (i.e., 0.45 counts per pixel in the vacuum region).

A.2   Experimental parameters

In phase-contrast TEM imaging (the technique employed here), multiple electron-optical and specimen parameters can give rise to complex, non-linear modulations of the image contrast. These parameters can include the objective lens defocus, the specimen thickness, the orientation of the specimen, and its crystallographic shape/structure. Due to the complex image formation mechanisms, atomic columns of the same material imaged may appear black or white (or somewhere in between, i.e., intermediate) depending on the exact combination of these various factors. Examples of the type of contrast reversal that may occur for a static structure imaged at constant thickness and tilt are given
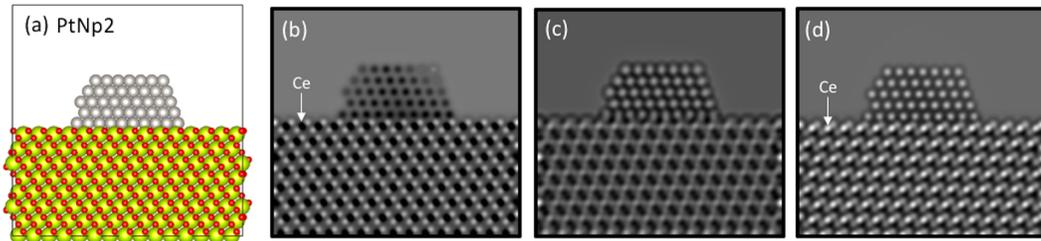
Figure 6: **Demonstration of contrast reversal with changes in defocus.** (a) Image of the Pt/CeO$_2$ atomic structural model.(b) to (d) Simulated images under different electron-optical focusing conditions, emphasizing variations on the Ce and Pt column contrast. In (b), the image shows black contrast for both Ce and Pt columns. In (c), the Pt columns reverse contrast and now appear white, while Ce columns become challenging to discriminate. Finally, in (d) all of the atomic columns appear with white contrast.
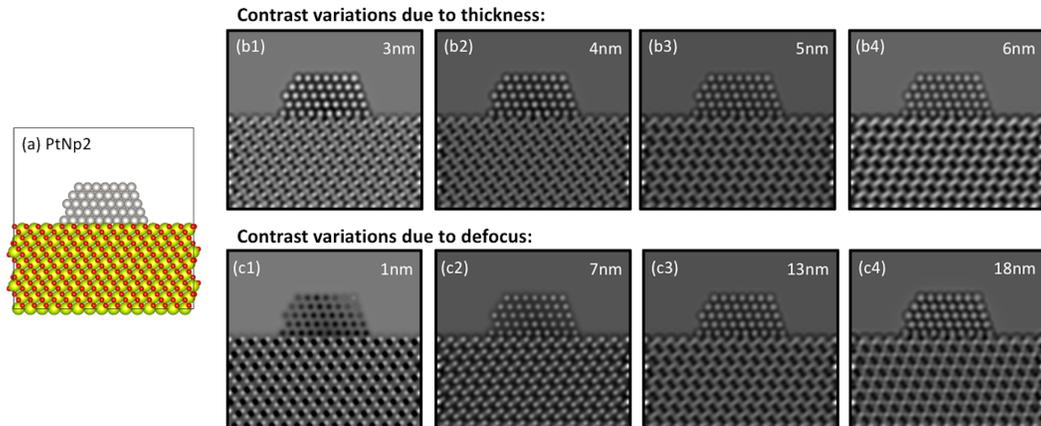

Figure 7: **Image contrast variations due to thickness and defocus.** (a) Image of the Pt/CeO$_2$ atomic structural model. (b1) to (b4) Simulated images at a defocus value of 13 nm, where the variations of the contrast are due to the thickness of the model, increased from 3 nm to 6nm. (c1) to (c4) Contrast variations on simulated images due to defocus: the thickness of the model has been kept constant at 5 nm and the defocus has been tuned to 1 nm, 7 nm, 13 nm, and 18 nm.

in Figure 6. Additionally, images showing the type of contrast variations that may occur when the support thickness is changed, and how these compare to those which arise from changes in defocus are given in Figure 7.

Here, we systematically varied these parameters to generate a large number of cases (approximately 18,000), corresponding to potential combinations that may arise during a real experiment (see Figure 8). First, around 100 atomic-scale structural models of CeO$_2$-supported Pt nanoparticles were generated. Each model represents Pt nanoparticles of various size, shape and atomic structure (e.g., small, medium, or large size, with either faceted or defected surfaces, or some combination of both), supported on CeO$_2$, which itself may present either a faceted surface or one characterized by surface defects. Secondly, the thickness of the CeO$_2$ support was varied from 3 nm to 6 nm along 1 nm increments. One aspect to note is that the thickness variation is not equally applied to each of the aforementioned models. Third, each resultant model was tilted from 0° to 4° about the x and y axes independently in increments of 1°. Thus, variations from 0° in x and 0° in y, to 4° in x and 0° in y, or 0° in x and 4° in y were considered. The final parameter systematically varied in the simulated image dataset was the electron optical defocus. Every model containing a unique shape/structure, thickness, and tilt (855 total) was imaged under a range of defocus values which often arise experimentally. Namely, the defocus was varied from 0 nm to 20 nm, along increments of 1 nm. Considering all combinations of the varied parameters, a total of 17,955 simulated images were generated for training and testing the neural network.
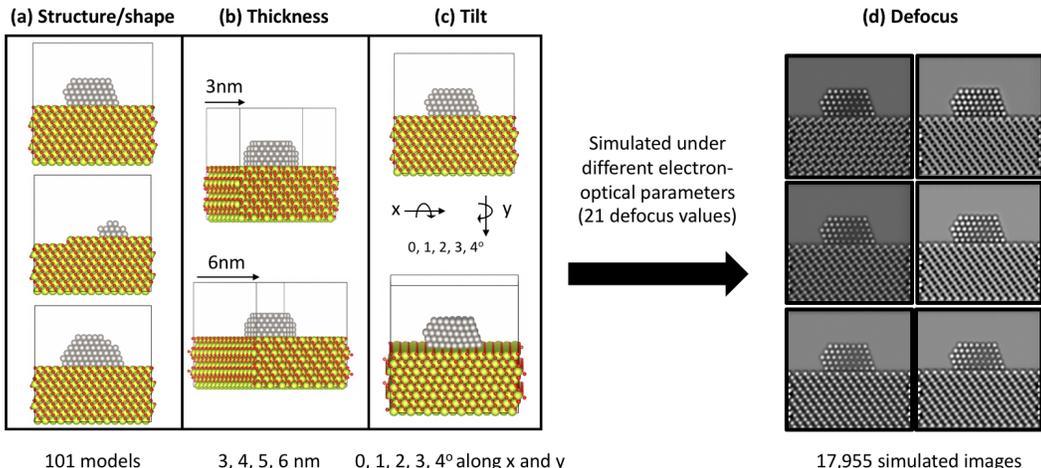
| (a) Structure/shape | (b) Thickness | (c) Tilt | (d) Defocus |

Figure 8: **Summary of parameters considered during the modelling and image simulation processes.** Subset of Pt/CeO$_2$ atomic models presenting variations on the (a) structure and shape of the nanoparticle and the support, (b) the thickness of the CeO$_2$ slab and (c) the tilt of the atomic models. Color code for the models matches Pt, Ce and O with grey, yellow and red atoms respectively. (d) Simulated images under different defocus values.

A.3   DESCRIPTION OF NANOPARTICLE STRUCTURES

The 3D atomic structural models utilized in this work consist of faceted Pt nanoparticles that oriented in a [110] zone axis and that are supported on a CeO$_2$ (111) surface which is itself oriented in the [110] zone axis. This crystallographic configuration corresponds to that which is often observed experimentally and is thus the focus of the current work. All of the models have been constructed with the freely available Rhodius software (Bernal et al., 1998). Each model consists of a supercell having x and y dimensions of 5 nm x 5 nm. As discussed above, the support thickness was systematically varied for each model, and so the supercell's z dimension varies between 3 nm and 6 nm, depending on the thickness of the particular model.

While imaging these materials systems, experimentalists often aspire to visualize atomic-level structural rearrangements that can occur at the surfaces of the supported nanoparticles. Additionally, there are many millions of nanoparticles on a typical TEM sample, and the specific atomic-scale structural features comprising the surfaces of those imaged during an experiment may vary slightly from nanoparticle to nanoparticle. In order to encompass such complexity in the training dataset, a variety of Pt nanoparticles of multiple sizes/shape and surface defect character were incorporated into the 3D models. For example, four such models of CeO$_2$-supported Pt nanoparticles having various size and shape are shown in parts (a) to (d) of Figure 9. The multi-slice TEM image simulations generated from the models are shown below each for two different conditions, namely in parts (a1) to (d1), images are shown for a case in which the support is 3 nm thick, the defocus is 9 nm, and the tilt is 0° in x and 0° in y; in parts (a2) to (d2), images are shown for the case in which the support is 5 nm thick, the defocus is 6 nm, and the tilt is 4° in x and 0° in y. Furthermore, the surface character of the Pt nanoparticles was varied by altering the defect structure at different surface sites. A few examples are depicted in Figure 10. Here, in part (a), a CeO$_2$-supported Pt nanoparticle with faceted surfaces is shown; directly beneath it in (a1) is an image simulated under conditions in which the support is 3 nm thick, the defocus is 9nm, and there is no tilt. The arrowed sites designate locations on the Pt surface that have been subsequently altered. In part (b), the surface has been modified by removing a full atomic column from the arrowed location. In part (c), the occupancy of the arrowed corner site has been reduced by half. And in part (d), the occupancy of the arrowed corner site has been further reduced to a single atom. Parts (b1) - (d1) show the images simulated from these respective structures under the same imaging condition. Note that the surface sites altered in the structure correspond to high-energy sites (e.g., corners and edges) which are more likely to dynamically rearrange or show variation than, say, a low-energy terrace site located in the middle of the surface.
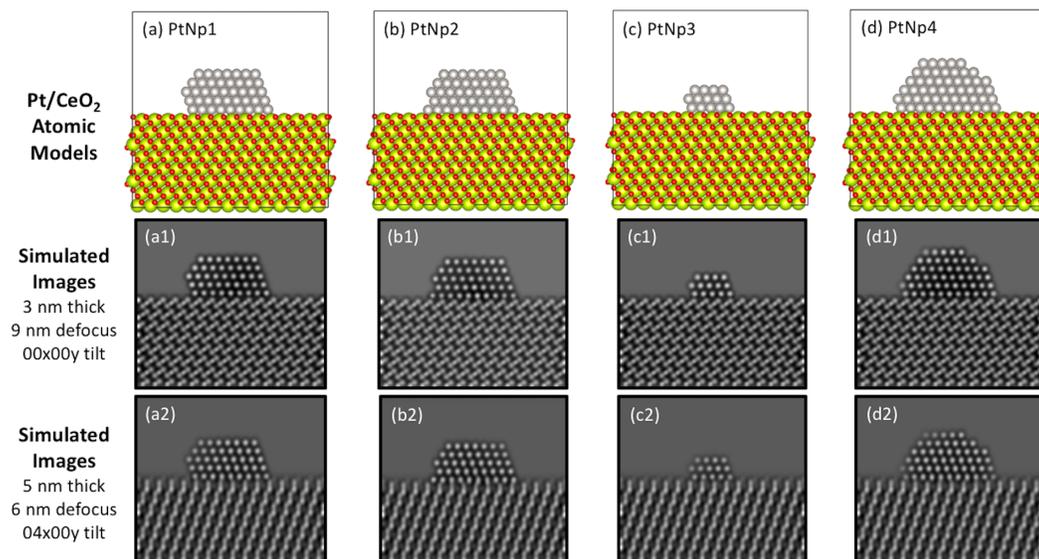
14

Figure 9: **Variations in the structure/size of the supported Pt nanoparticle.**(a) to (d) Atomic models of Pt nanoparticles (grey atoms) with different shapes and supported over a CeO_2 slab (yellow and red atoms respectively). (a1) to (d1) Simulated images depicting the described atomic models, considering a thickness of 3 nm, 9 nm of defocus and no tilt on the model, whereas (a2) to (d2) illustrate the same model under different conditions: 5 nm thickness, 6 nm defocus and 4 degrees tilted along x axis. All the simulated images present a Cs value of -13 μm.
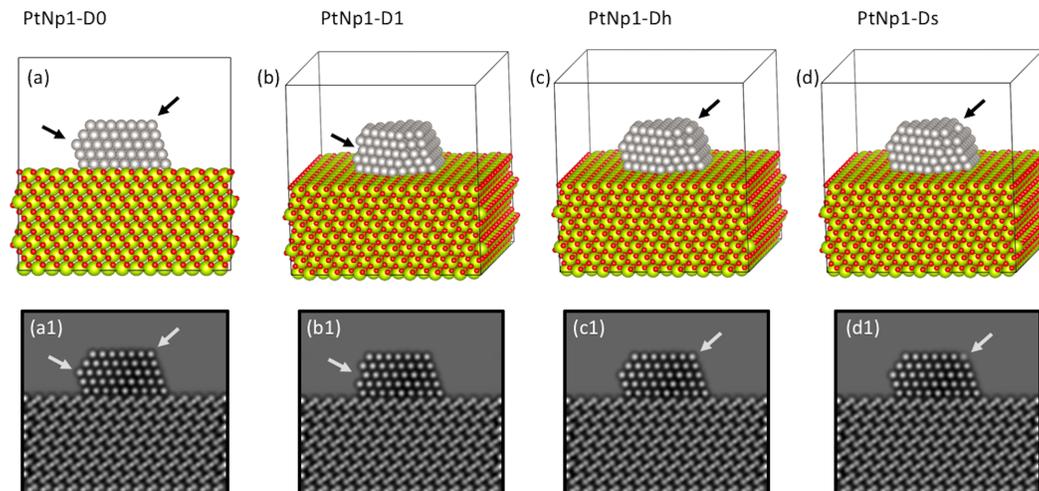

Figure 10: **Variations in the defects of the Pt surface structure.** (a) Atomic model of a CeO_2-supported Pt nanoparticle without any defects. The surface has been modified by (b) removing a full atomic column, (c) removing half of the occupancy and (d) keeping a single atom. Black arrows point the sites where these defects are taking place. Models (b), (c) and (d) have been slightly tilted to observe these modifications. (a1) to (d1) Simulated images of the presented atomic columns considering a 3nm thickness, 9 nm defocus and no tilt.
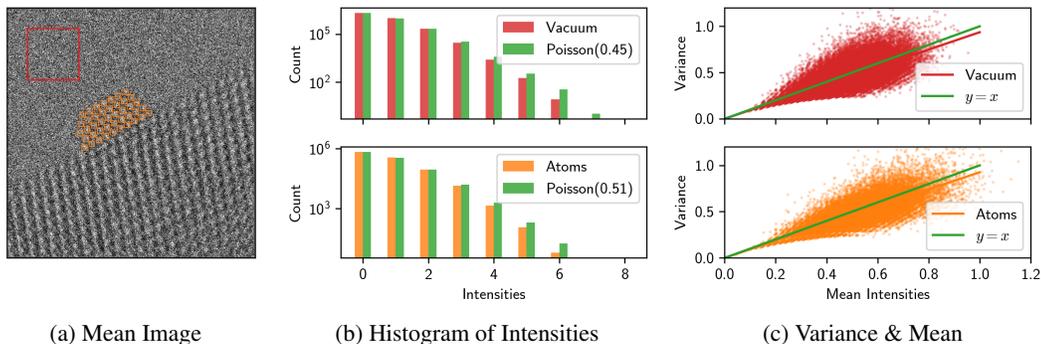
(a) Mean Image   (b) Histogram of Intensities   (c) Variance & Mean

Figure 11: **Analysis of the noise in the real data**. The analysis shows that the noise is approximately iid Poisson. (a) Pixel-wise mean over 40 frames of the real data described in Section 4. (b) Histogram of noisy pixel intensities from highlighted regions in the image matches, compared to a simulated Poisson distribution. (c) The plot of empirical mean and standard deviation of pixels approximately follows a line with unit slope, as expected from iid Poisson samples (the spread is due to averaging over only 40 frames).

B   Noise model

Given the physical origin of the noise in the experimental image acquisition process, we expect the noise to be dominated by shot noise, which can be modeled with a Poisson distribution. Here, the images were acquired on a direct electron detector operating in electron counting mode. In such conditions, the electron dose rate per pixel is sufficiently low enough that individual electron arrivals can be detected and registered. It is well known that the statistical fluctuations of such counting processes for discrete events are governed by shot noise. Additionally, we expect that other sources of noise, including fixed pattern noise, dark noise, and thermal noise are minimal after applying a gain correction and a dark reference to the raw image, and by cooling the detector to -20 °C, respectively. Readout noise is considered negligible since the pixels are read out individually. Thus, we expect that the noise in the counted TEM micrographs can be modelled as Poisson.

We empirically verified that the noise indeed follows Poisson statistics through the analysis in Figure 11. Our real dataset, described in Section 4 consists of 40 noisy frames acquired sequentially across time in 0.025 sec intervals. Figure 11(a) shows the mean image over all 40 frames. The region of the image containing no material (e.g., red box) corresponds to the vacuum, where the electron beam intensity is uniform. Fluctuations of the intensity in this region therefore purely arise from the noise. We validate that the histogram of these pixels aggregated over a spatial region that is approximately constant closely follows a Poisson distribution (Figure 11(b)). Pixels aggregated over spatial domains corresponding to the Pt atomic columns (e.g., orange boxes) show similar behavior. The empirical mean and variance, computed by averaging over the 40 time frames at every spatial location, follows a linear trend, further confirming that the noise distribution has Poisson properties (Figure 11(c)). We note that the spread in the scatter-plot in Figure 11(c) is due to the limited number of time frames over which we average.

C   Description of denoising CNNs

In this section we describe the CNN architectures and training procedure used for our computational experiments in more detail.

C.1   Proposed Architecture: UNet with large field of view

We propose to use a modified version of UNet Ronneberger et al. (2015) with 4 scales to achieve a large field of view. The network consisting of 4 `down-blocks` and 4 `up-blocks`. A `down-block` consists of a max-pooling layer, which reduces the spatial-dimension by half, followed by a `conv-block`. Similarly, an `up-block` consists of bilinear upsampling, which enlarges the size of the feature-map by a factor of two, followed by `conv-block`. A `conv-block` consists of

conv-BN-ReLU-conv-BN-ReLU, where conv represents a convolutional layer and BN stands for batch normalization (Ioffe & Szegedy, 2015). In our final model, we use 128 channels in each layer of `conv-block`.

C.2   DNCNN

DnCNN (Zhang et al., 2017) consists of 20 convolutional layers, each consisting of 3×3 filters and 64 channels, batch normalization (Ioffe & Szegedy, 2015), and a ReLU nonlinearity. It has a skip connection from the initial layer to the final layer, which has no nonlinear units.

C.3   SMALL UNET FROM DURR

We use the UNet proposed for the restoration module in DURR Zhang et al. (2018). The architecture consists of the following:

1. *conv1* - Takes in input image and maps to 32 channels with 5×5 convolutional kernels.
2. *conv2* - Input: 32 channels. Output: 32 channels. 3×3 convolutional kernels.
3. *conv3* - Input: 32 channels. Output: 64 channels. 3×3 convolutional kernels with stride 2.
4. *conv4*- Input: 64 channels. Output: 64 channels. 3×3 convolutional kernels.
5. *conv5*- Input: 64 channels. Output: 64 channels. 3×3 convolutional kernels with dilation factor of 2.
6. *conv6*- Input: 64 channels. Output: 64 channels. 3×3 convolutional kernels with dilation factor of 4.
7. *conv7*- Transpose Convolution layer. Input: 64 channels. Output: 64 channels. 4×4 filters with stride 2.
8. *conv8*- Input: 96 channels. Output: 64 channels. 3×3 convolutional kernels. The input to this layer is the concatenation of the outputs of layer *conv7* and *conv2*.
9. *conv9*- Input: 32 channels. Output: 1 channels. 5×5 convolutional kernels.

This configuration of UNet assumes even width and height, so we remove one row or column from images in with odd height or width.

C.4   TRAINING PROCEDURE

All CNNs were trained on 400×400 patches extracted from the training images and augmented with horizontal flipping, vertical flipping, random rotations between $-45°$ and $+45°$ and random resizing by a factor of 0.75-0.82. The models were trained using the Adam optimizer (Kingma & Ba, 2014), with a default starting learning rate of 10^{-3}, which was reduced by a factor of 2 every time the validation PSNR plateaued. Training was terminated via early stopping based on validation PSNR.

D   DESCRIPTION OF BASELINE METHODS

Low-pass Filter (LPF): The cut-off frequency of a linear low-pass filter was set in order to preserve the signal information, while suppressing the noise.

Adaptive Wiener Filter: As Poisson noise is signal dependent, we use an adaptive low-pass Wiener filter (Lim, 1990) to perform smoothing with local mean and variance around each pixel estimated from its local neighborhood. We performed a grid search over neighborhood sizes of $3, 13, 27$ for the filter, selecting 13 based on expert evaluation of the denoised images in the validation set.

VST + BM3D and VST + non-local means (NLM): BM3D (Makitalo & Foi, 2012) and NLM (Buades et al., 2005) are popular denoising techniques for natural images with additive Gaussian noise. Following (Zhang et al., 2019), we use a nonlinear variance-stabilizing transformation (VST) to convert the Poisson denoising problem into a Gaussian noise removal problem. Empirically, the experimental pixel intensities follow Poisson distributions (Figure 11), so we apply the Anscombe transformation that is designed for Poisson-Gaussian noise on the noisy image, apply BM3D or NLM

MODEL	Parameters	Receptive Field	PSNR	SSIM
SBD + DnCNN (Zhang et al., 2017)	668K	41×41	30.47 ± 0.64	0.93 ± 0.01
SBD + Small UNet (Zhang et al., 2018)	233K	45×45	30.87 ± 0.56	0.93 ± 0.01
SBD + UNet (32 base channels)	352K	221×221	36.39 ± 0.77	0.98 ± 0.01
SBD + UNet (64 base channels)	1.41M	221×221	37.24 ± 0.76	0.99 ± 0.01
SBD + UNet (128 base channels)	5.61M	221×221	$\mathbf{38.05} \pm 0.81$	$\mathbf{0.99} \pm 0.01$

Table 2: **Comparison of different CNN architectures.** Mean PSNR and SSIM ($\pm$ standard deviation) of different CNN architectures on the held-out simulated test set described in Section 5. Note that increasing the field of view of the UNet from 45×45 pixels to 221×221 produces a dramatic increase of around 6 dB in PSNR, even if the number of parameters remains similar. Further increasing the number of parameters produces a modest gain in performance.

to the transformed image, and finally use the exact unbiased inverse transformation to recover the denoised image (Makitalo & Foi, 2012).

PURE-LET: PURE-LET (Luisier et al., 2010) is a transform-domain thresholding algorithm adapted to mixed Poisson–Gaussian noise. The method requires the input image to have dimensions of the form $(2^k, 2^k)$. To apply this method to our test images, we extracted 128×128 overlapping patches from the image, denoised individual patches and stitched them back together by averaging over the overlapping pixels.

Blind-spot denoising: We trained a blind-spot network based on UNet (Laine et al., 2019) on the real TEM dataset. Following (Laine et al., 2019) we trained the network on patches obtained from the images. We used 600×600 and used Adam optimizer (Kingma & Ba, 2014) with a starting learning rate of 1×10^{-4} which was reduced by in half for every 2000 epochs. We trained for 5000 epochs.

E ADDITIONAL RESULTS

In this section we include the following additional results:

- Figure 12 shows that CNNs may easily overfit the orientation and scaling of the images in the training dataset.

- Figure 13 shows that the network is robust to variations in imaging parameters and signal structure. An exception is the network trained only on black contrast which does not generalize well to other imaging contrasts.

- Figure 14, 15 shows two examples of simulated images denoised using the proposed approach and the methods described in Sections C and D.

- Figure 16, 17 shows two examples of real images denoised using the proposed approach and the methods described in Sections C and D.

- Figure 18 shows a gradient analysis of the learned denoising function, which reveals that the network learns to combine local and non-local information to estimate the denoised image.
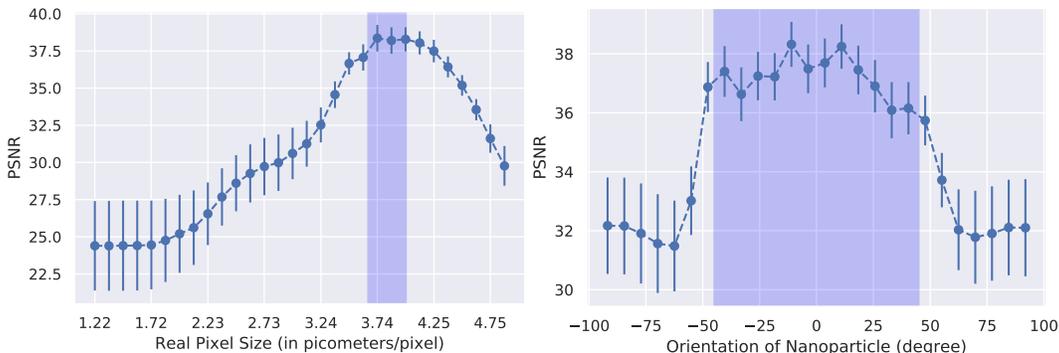
Figure 12: **Overfitting scaling and orientation**. The plots show the performance of the proposed network in PSNR for simulated images that have different scalings (on the left, measured in terms of the corresponding pixel size in picometers) and orientations (on the right). The CNN was trained on data augmented with rescaled and rotated images corresponding to the regions shaded in purple. When tested out of those regions, the denoising performance is degraded significantly. This shows that careful data augmentation is required to ensure invariance to scaling and orientation.

(a) Contrasts (b) Structures (c) Defects
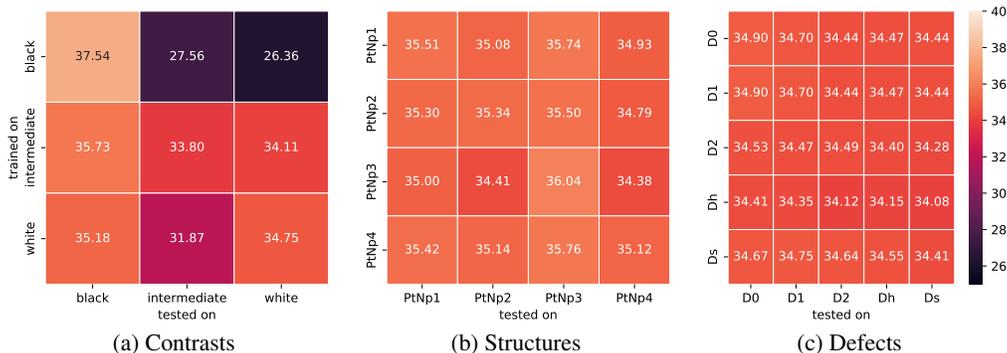
Figure 13: **Generalization across different imaging parameters and signal structures**. In order to study the generalization ability of the proposed method we divided the simulated dataset described in Section 4 into subsets, based on the atomic column contrast, (2) the structure/size of the supported Pt nanoparticle, and (3) the defects of the Pt surface structure. The tables show the test PSNR for networks trained and tested on different combinations of the subsets. (a) Networks trained on white and intermediate contrast generalize well to all other contrasts. The network trained on black contrast does not generalize to the others. (b) Networks trained on a type of nanoparticle structure generalize well to all other types. (c) Networks trained on one type of defect or no defects generalize well to different types. Please refer to Fig 6, 10, 9 and 7 to see the effect of these parameters on the images.
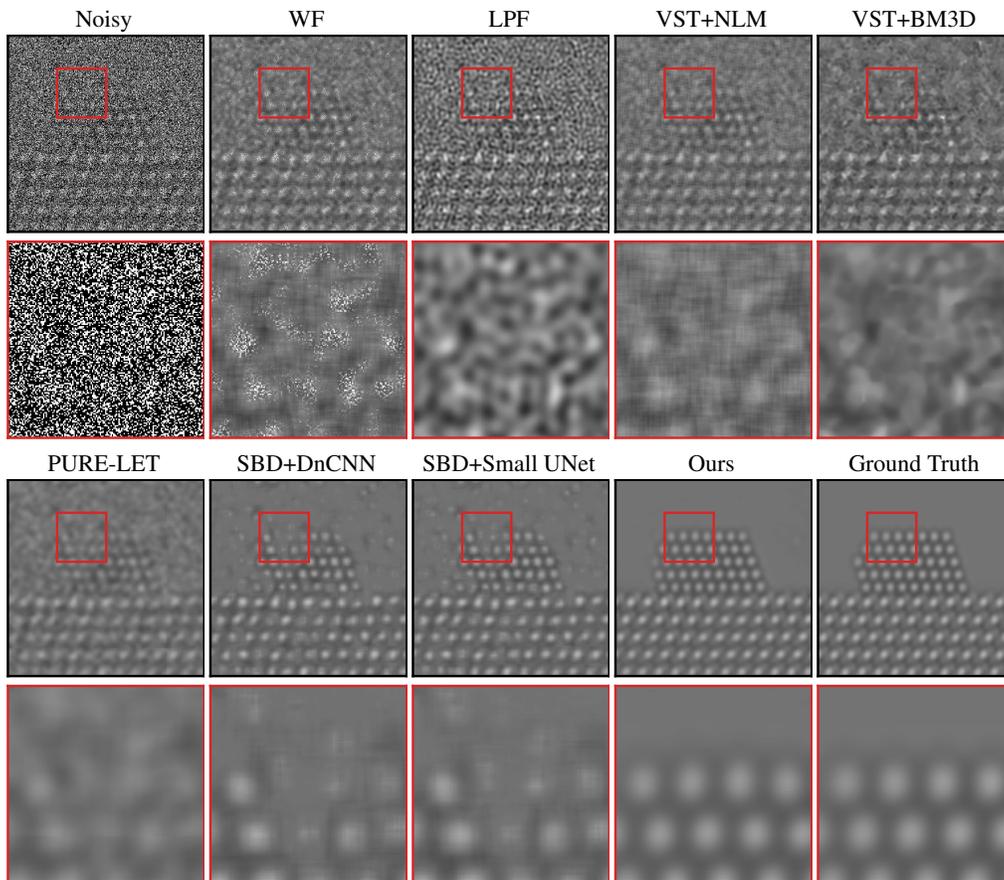
Figure 14: **Denoising results for simulated data**. Our proposed approach produces denoised images of much higher quality than the other approaches, described in Sections C and D. In contrast to the other methods, SBD combined with the proposed architecture is able to precisely recover the structure of the nanoparticle and has very few artifacts.
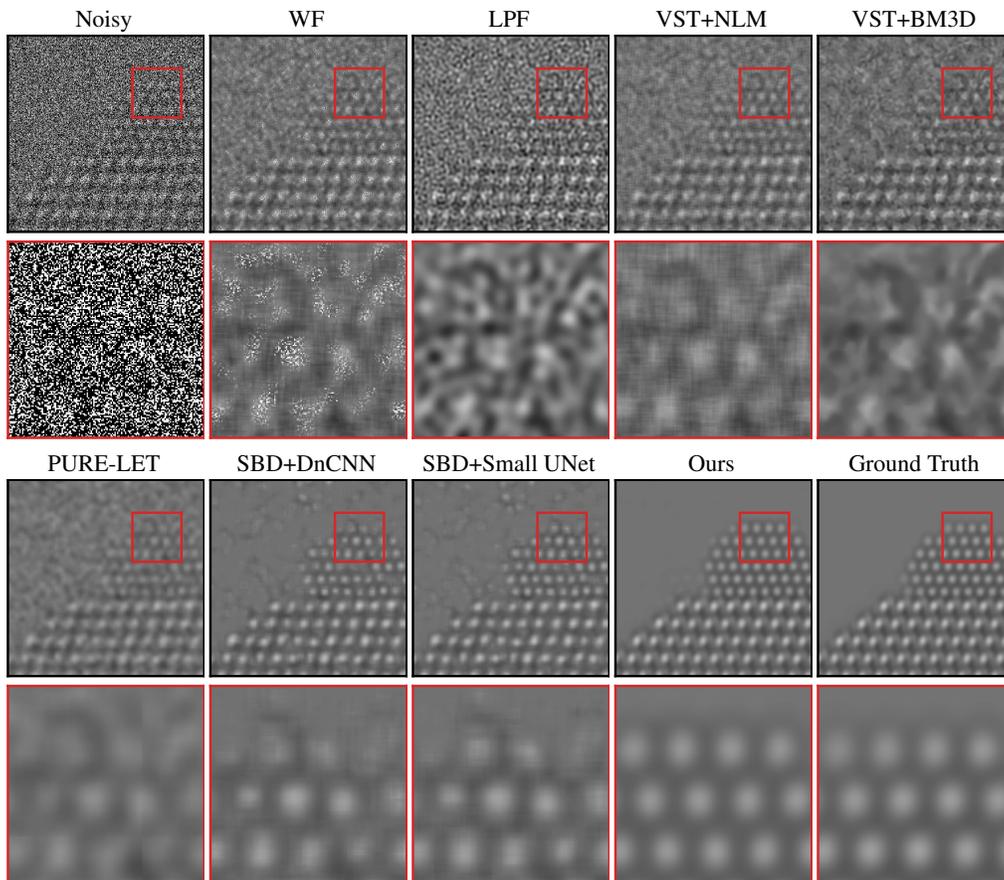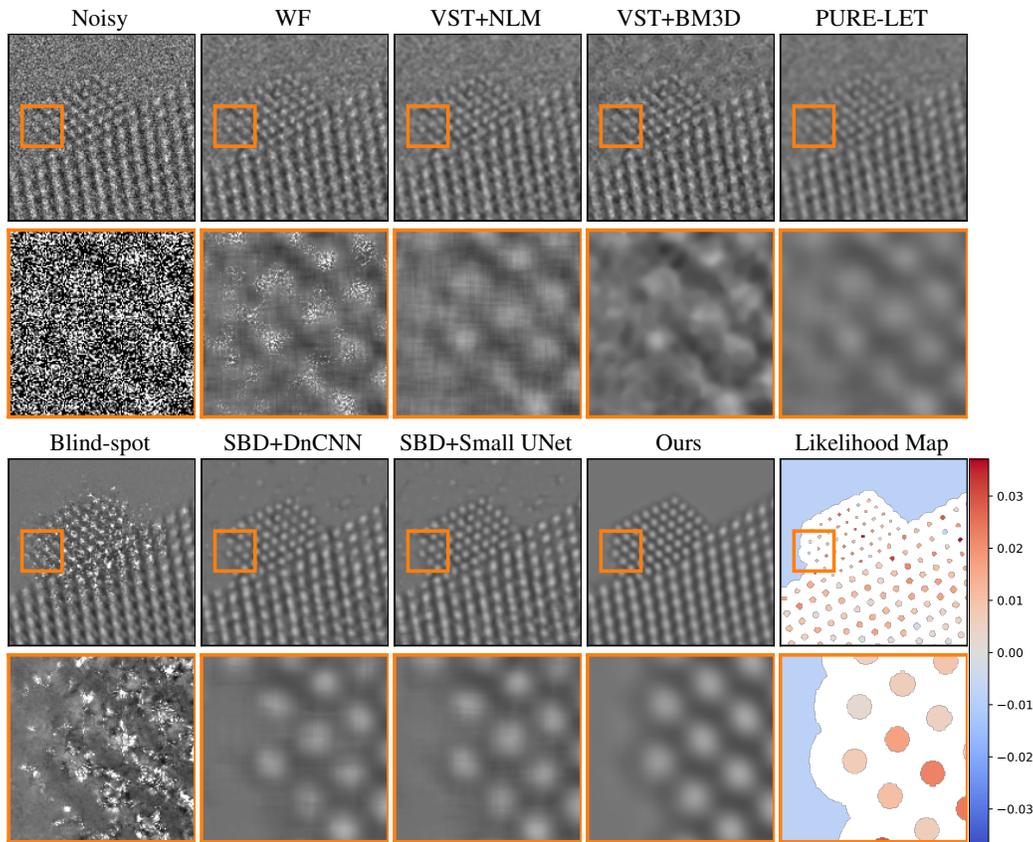
Figure 15: **Denoising results for simulated data**. Our proposed approach produces denoised images of much higher quality than the other approaches, described in Sections C and D. In contrast to the other methods, SBD combined with the proposed architecture is able to precisely recover the structure of the nanoparticle and has very few artifacts.

Figure 16: **Denoising results for real data**. Our proposed approach produces denoised images of much higher quality than the other approaches, described in Sections C and D. In contrast to the other methods, SBD combined with the proposed architecture is able to precisely recover the structure of the nanoparticle and has very few artifacts in the vacuum region. The likelihood map quantifies the agreement between the observed data and the detected regions corresponding to atomic columns and to the vacuum.
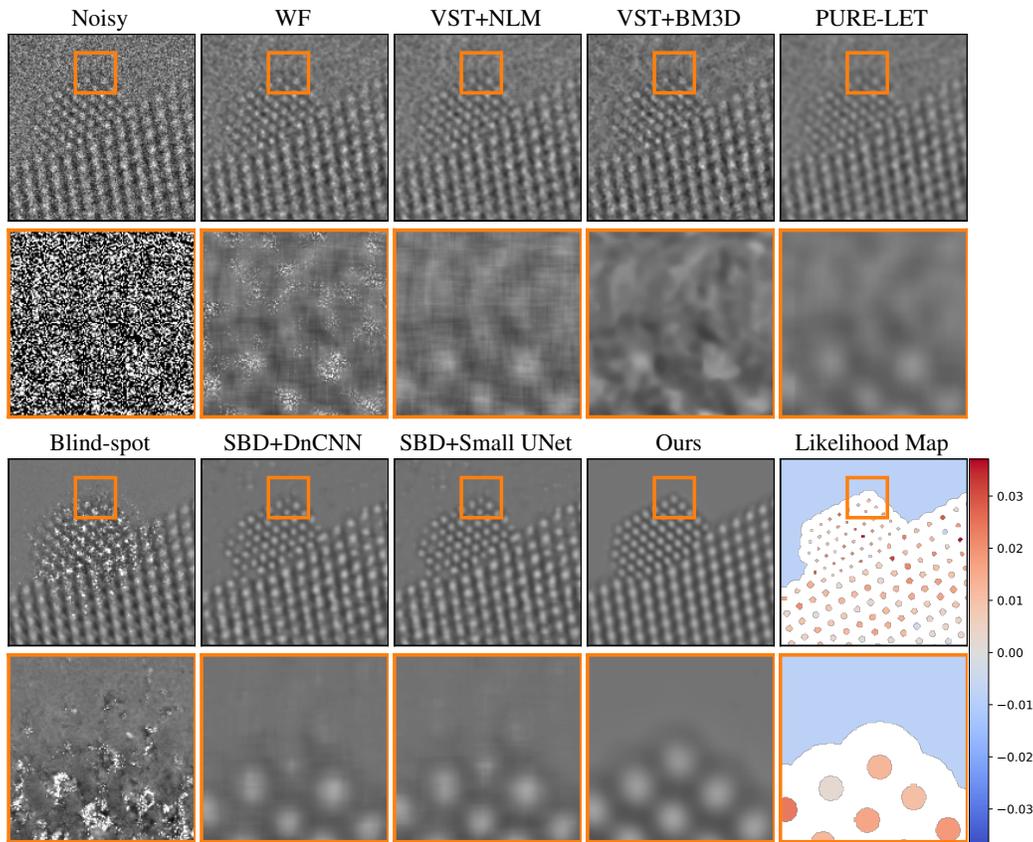
Figure 17: **Denoising results for real data**. Our proposed approach produces denoised images of much higher quality than the other approaches, described in Sections C and D. In contrast to the other methods, SBD combined with the proposed architecture is able to precisely recover the structure of the nanoparticle and has very few artifacts in the vacuum region. The likelihood map quantifies the agreement between the observed data and the detected regions corresponding to atomic columns and to the vacuum.
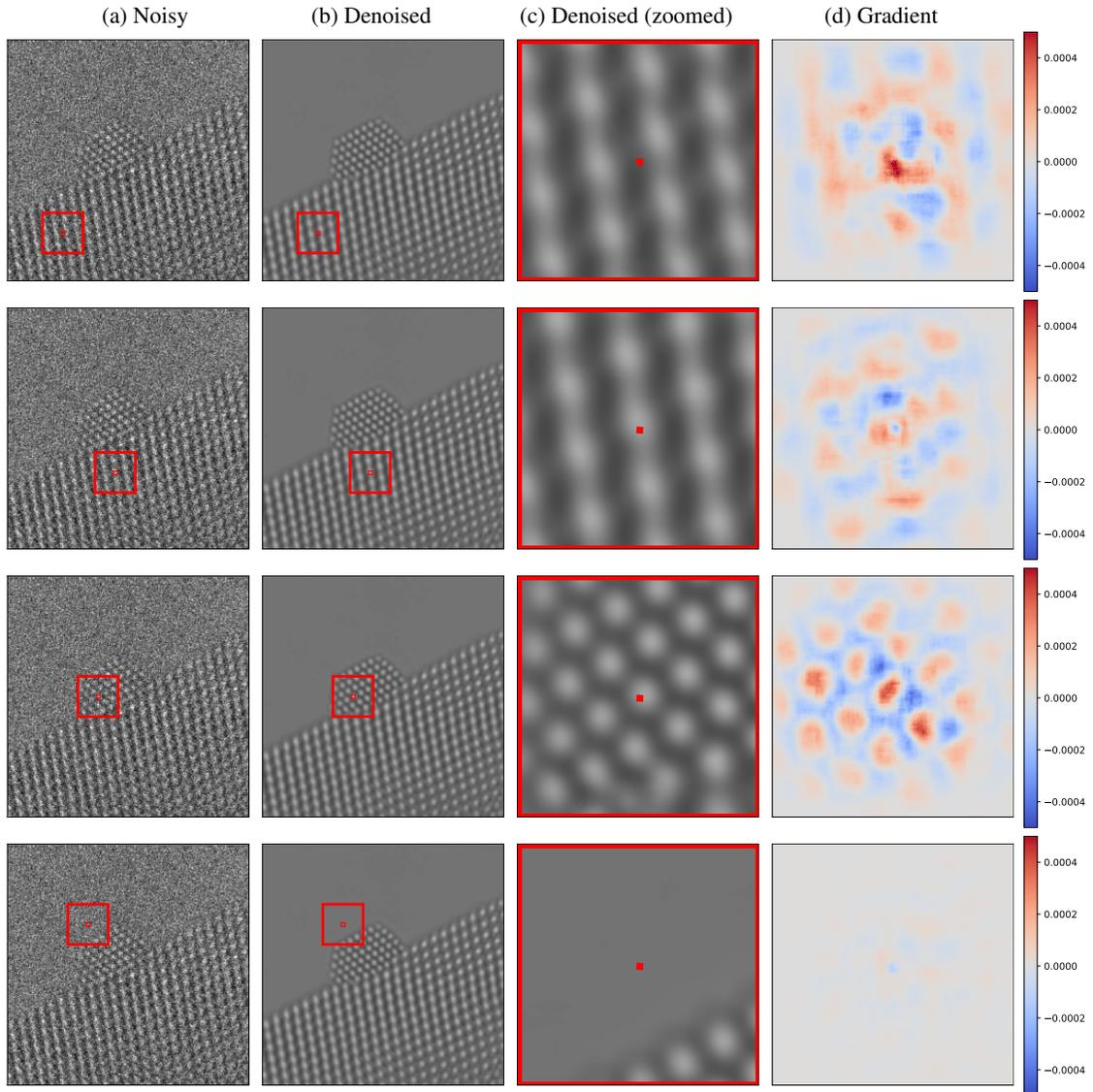
Figure 18: **Gradient analysis of the learned denoising function on real data.** (a) To compute the red pixel in the denoised image (b), the proposed CNN uses a 220×220 area (red box) in the noisy image (a). The gradient of the denoised pixel with respect to its input indicates what regions in the noisy image have a greater influence on the estimate (according to a first-order Taylor approximation to the denoising map). The gradient (d) weights nearby pixels more heavily, but also has significant magnitude at pixels located on different atoms. This suggests that the CNN combines local and non-local information to estimate the pixel. The colorbar is shared across all the images.