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

Sreyas Mohan, Ramon Manzorro, Joshua L. Vincent, Binh Tang, Dev Y. Sheth, Eero P. Simoncelli, *Fellow, IEEE*, David S. Matteson, Peter A. Crozier, and Carlos Fernandez-Granda

***Abstract*—Denoising is a fundamental challenge in scientific imaging. Deep convolutional neural networks (CNNs) provide the current state of the art in denoising photographic images. However, their potential has been inadequately explored for scientific imaging. Denoising CNNs are typically trained on clean images corrupted with artificial noise, but in scientific applications, noiseless ground-truth images are usually not available. To address this, we propose a simulation-based denoising (SBD) framework, in which CNNs are trained on simulated images. We test the framework on transmission electron microscopy (TEM) data, showing that it outperforms existing techniques on a simulated benchmark dataset, and on real data. 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. Our results reveal that state-of-the-art architectures for denoising photographic images may not be well adapted to scientific-imaging data. For instance, substantially increasing their field-of-view dramatically improves their performance on TEM images acquired at low signal-to-noise ratios. We also demonstrate that standard performance metrics for photographs (such as peak signal-to-noise ratio) may not be scientifically meaningful, and propose several metrics to remedy this issue in the case of TEM images. In addition, we propose a technique, based on likelihood computations, to visualize the agreement between the structure of the denoised images and the observed data. Finally, we release a publicly available benchmark dataset containing 18,000 simulated TEM images.**

***Index Terms*—Denoising, scientific imaging, electron microscopy, deep learning.**

Manuscript received 2 February 2022; revised 21 April 2022; accepted 1 May 2022. Date of publication 20 May 2022; date of current version 20 July 2022. The work of Sreyas Mohan, who designed, implemented, and analyzed the proposed methodology, was supported in part by the NSF under NRT HDR Award 1922658. The work of Ramon Manzorro and Peter A. Crozier, who created the training dataset, was supported by NSF under Grant OAC-1940263. The work of Joshua L. Vincent and Peter A. Crozier, who acquired and processed the experimental data, was supported by the NSF under Grant CBET-1604971. The work of Binh Tang, who contributed to the implementation of methodology, and David S. Matteson, who supervised the design of methodology, was supported by NSF under Grant OAC-1940124 and Grant CCF-1934985. The work of Eero P Simoncelli, who supervised the design and analysis of methodology, was supported by Howard HughesMedical Institute (HHMI). The work of Carlos Fernandez-Granda, who supervised the design, implementation, and analysis of methodology, was supported by NSF under Grant OAC-1940097. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bo Zhao. *(Corresponding author: Carlos Fernandez-Granda.)*

Sreyas Mohan is with New York University, New York, NY 10012 USA (e-mail: sm7582@nyu.edu).

Ramon Manzorro, Joshua L. Vincent, and Peter A. Crozier are with the Arizona State University, Tempe, AZ, USA (e-mail: ramon.manzorro@gmail.com; joshua.vincent@asu.edu; crozier@asu.edu).

Binh Tang and David S. Matteson are with the Cornell University, Ithaca, NY 14850 USA (e-mail: bvt5@cornell.edu; matteson@cornell.edu).

Dev Y. Sheth is with the Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai 600036, India (e-mail: cs17b106@cse.iitm.ac.in).

Eero P. Simoncelli is with the CNS/CIMS/CDS, New York University, New York, NY 10001 USA, and also with Flatiron Institute, New York, NY 10010 USA (e-mail: eero.simoncelli@nyu.edu).

Carlos Fernandez-Granda is with the Department of Mathematics, Courant Institute of Mathematical Science, and Center for Data Science, New York University, New York, NY 10003 USA (e-mail: cfgranda@cims.nyu.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TCI.2022.3176536, provided by the authors.

Digital Object Identifier 10.1109/TCI.2022.3176536

I. Introduction

Imaging technology is an essential tool in many scientific domains. Electron microscopy enables the visualization of atomic structures [1], fluorescence microscopy makes it possible to study cellular processes [2], and telescopes reveal galaxies and other astronomical objects that are light years away [3]. In all these modalities, images are corrupted by noise associated with stochastic processes occurring during signal generation and detection, degrading the information content of the image data. The general goal of denoising is to estimate and restore the information missing from these noisy observations, thus facilitating the extraction of useful scientific information.

In the past decade, convolutional neural networks (CNNs) [4] have achieved state-of-the-art performance in image denoising [5], [6]. However, the potential of this methodology has barely been explored in the context of scientific imaging. In the vast majority of the existing work, noisy data are generated by adding Gaussian noise to clean photographs. The CNNs are then trained to approximate the ground-truth images from these measurements, usually by minimizing mean squared error [5]. Unfortunately, *this paradigm is not adequate for most scientific domains*, where large, labeled datasets of ground-truth clean data are typically not available. 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 [8], [9]. The technique has had a huge impact in structural biology, as recognized with the award of the 2017 Nobel Prize in Chemistry [10]. Recent advancements in direct electron detection systems enable experimentalists to image dynamic events at frame rates in the kilohertz range [11], [12]. Imaging at these time scales is

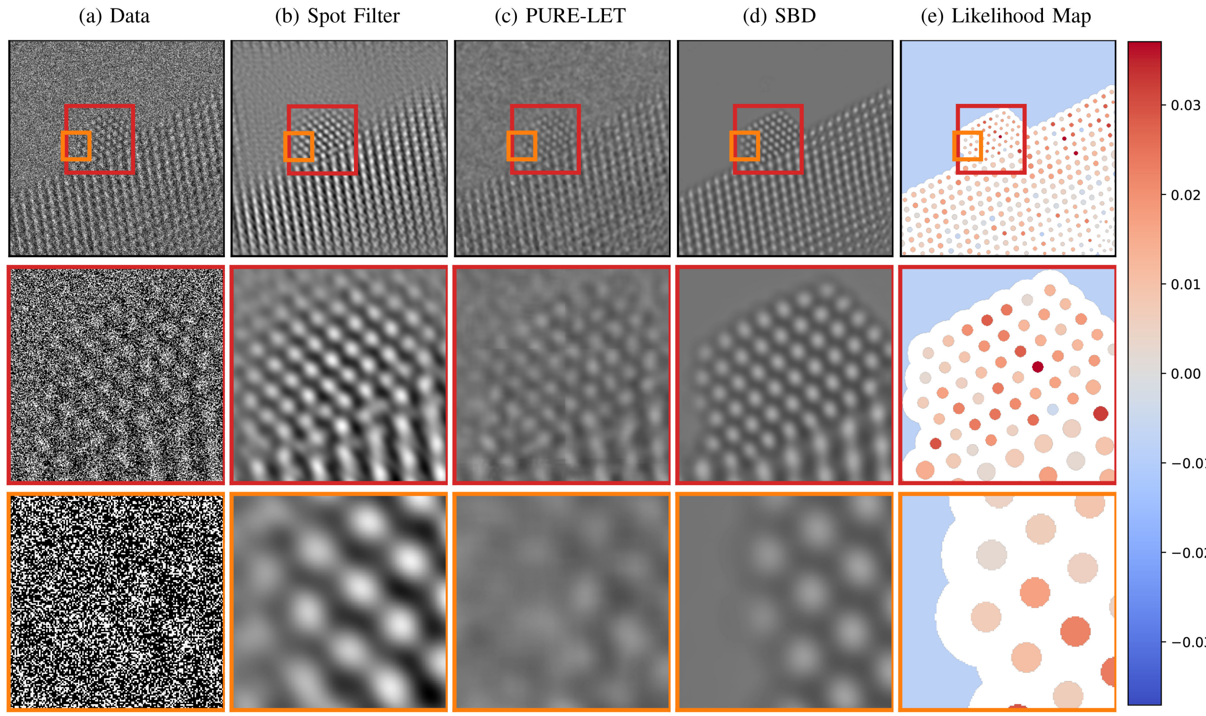

Fig. 1. Denoising results for real data. (a) An experimentally-acquired atomic-resolution transmission electron microscope image of a CeO2-supported Pt nanoparticle. A description of the experimental data acquisition is given in Section IV. The average image intensity is 0.45 electrons/pixel (i.e., a large fraction of pixels register zero electrons), which results in an extremely low signal-to-noise ratio. (b) Denoised image obtained via Fourier-based filtering by a domain expert (see Fig. SM10 for the mask used). (c) Denoised image obtained via the wavelet-based PURE-LET method [7]. (d) Denoised image obtained by the proposed simulation-based denoising (SBD) framework. (e) Likelihood map quantifying to what extent the atomic structure identified from the SBD denoised image is consistent with the data (see Section III). 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.

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 [13]–[17]. 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 by increasing the intensity of the incident electron beam, since the high-energy beam can also damage the material when exposed to high doses. Consequently, there is an acute need for novel denoising technology in this domain.

In order to apply the proposed SBD framework to TEM data, we generate a simulated dataset of TEM images, containing 18,000 examples, and use it to train CNNs for noise removal. This approach outperforms existing techniques by a wide margin on held-out[1] simulated data, as well as on real TEM measurements (see Figs. 1, SM12 and SM16 and Sections V-B and V-D). We perform a thorough analysis of the generalization capability of our models, demonstrating that the CNNs are robust to variations of imaging parameters and of underlying signal structure. Our results indicate that architectures optimized for natural photographic images may have fundamental shortcomings when applied to domain-specific data. For instance, we show that substantially increasing the field-of-view of denoising CNNs has almost no effect on photographs, but produces a significant boost in performance for TEM images. We also demonstrate that standard performance metrics for photographs, such as peak signal-to-noise ratio (PSNR) (Section SM ii) and structural similarity index (SSIM) [18], often fail to produce a scientifically-meaningful evaluation of the denoising results. For example, the presence or absence of a single atomic column often results in a negligible change in these metrics. This is highly problematic, because detecting these columns is one of the main motivations for our case study. To remedy this issue, we propose several scientifically-motivated metrics to evaluate our results (see Section V-C). In addition, we propose a likelihood-based visualization of the agreement between the observed measurements and structures of interest (such as atomic columns) in the denoised image. This visualization can be used to flag denoising artefacts, which may be mistaken for scientifically-relevant structure (see Fig. 3). Finally, to encourage further development of deep-learning methodologies for scientific imaging, we have made our benchmark dataset of TEM images publicly at https://sreyas-mohan.github.io/electron-microscopy-denoising/. More details on applying the proposed methodology to TEM data, and domain-specific insights derived from the denoised images are described in our companion paper [19].

[1]part of the dataset not used for training.

II. Related Work

1) Denoising in scientific imaging: A wide variety of denoising methods have been applied across different scientific imaging modalities, including traditional linear filters [20], nonlinear filters [21]–[23], wavelet-based methods [24]–[27], and sparsity-based approaches [27], [28]. Deep convolutional networks have been shown to outperform all of these approaches in photographic images [5], [6]. 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, CNN-based denoising has been applied to low-dose computer tomography [29], positron-emission tomography [30] and scintillation-camera data [31]. Refs. [32], [33] apply CNNs to denoise simulated electron microscopy data, without validating on real data. Ref. [34] train CNN to denoise by adding synthetic noise on high-quality electron micrograph data. Ref. [35] trains CNNs to denoise Raman scattering microscopy 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 challenge of gathering adequate datasets to train the deep networks. In this work, we propose to address this challenge by training denoising CNNs on carefully-designed simulated datasets, and validate our approach on experimental measurements.

2) Unsupervised denoising: Unsupervised denoising is a promising approach for applications where ground-truth images are not available. Unsupervised methods based on wavelets have achieved performance comparable to their supervised counterparts on photographic images [36]–[38].

Noise2Noise [39], 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 [40]. More recent methods can be trained directly on noisy images [41]–[43]. Several recent works apply this approach to fluorescence microscopy data [44]–[47]. In the case of our TEM data, standard unsupervised methods do not perform as well as the proposed supervised approach (see Section V-D1). This is possibly due to the limited number of training data (see Fig. SM18) and the low input SNR. The SNR of our TEM data (around 3 dB) is orders of magnitude lower than that reported in typical unsupervised denoising works (e.g. around 27 dB for [47]).

3) Deep Learning for TEM: Deep CNNs have been applied to other image-processing tasks in TEM beyond denoising, see [48] for a comprehensive review. Ref. [49] proposes a CNN-based method for TEM image super-resolution, wherein CNNs are trained on pairs of low-resolution and high-resolution images acquired experimentally. Ref. [50] applies CNNs to perform segmentation and systematically studies the influence of the design of the training dataset and network architecture on the generalization capabilities of these models. In this work, we provide a similar analysis for denoising. Refs. [51], [52], [53], and [54] propose a CNN-based method to identify structures of interest in TEM images. They train on carefully designed simulated data and show that the model generalizes to real data. Our work provides further evidence that CNNs trained on simulated data can generalize effectively to real measurements.

III. Methodology

A. Simulation-Based Denoising

Current state-of-the-art deep-learning techniques for denoising photographic images require a training set of ground-truth images [5]. Typically these clean images are corrupted with additive Gaussian noise, and the CNNs are trained to minimize 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).

Simulation-based denoising (SBD) consists of three stages: simulation of the training set, training of the CNNs using the simulated data, and inference on the real data (see Fig. 2 for an overview of the methodology). In order to generate the training set, we simulate clean images x_1, ..., $x_N \in \mathbb{R}^M$ (where M is the number of pixels) according to a predefined physical model. These clean images are then corrupted using a noise model, which can follow a predefined model or be learned from the data, to generate the simulated noisy data. We provide a detailed account of how we generate the simulated dataset for our case study in Sections IV and SM i-A, and of the noise model in Section SM iii. Let $Y(x_i)$ denote the random vector representing the noisy image corresponding to the clean simulated image x_i and let $y(x_i)$ represent a realization of $Y(x_i)$. We parameterize the denoising function as a CNN $f_\theta : \mathbb{R}^M \to \mathbb{R}^M$ where the parameters θ are the weights of the network. To find a good denoising function f_θ, we minimize a loss function $\mathcal{L} : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ which quantifies how close the estimate from the CNN $f_\theta(y(x_i))$ is to the clean image x_i. In our case study, we use mean squared error, which is a standard choice in CNN-based denoising [5]. More concretely, during the training stage, we compute the parameters by solving

$$\hat{\theta} = \arg\min_{\theta} \mathbb{E}\left[\sum_{i=1}^{N} \mathcal{L}(f_\theta(Y(x_i)), x_i)\right] \tag{1}$$

$$= \arg\min_{\theta} \mathbb{E}\left[\sum_{i=1}^{N} \|f_\theta(Y(x_i)) - x_i\|_2^2\right] \tag{2}$$

We perform minimization iteratively using a variant of stochastic gradient descent. We approximate the expectation in (1) by sampling new realizations of the noisy image $Y(x_i)$ every time we compute the gradient. Once the network is trained, it can be directly applied to new noisy images to perform denoising.

A crucial difference between SBD and previous methodology for deep denoising is that the training set needs to be explicitly *designed*. In order to ensure effective generalization to our real experimental 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

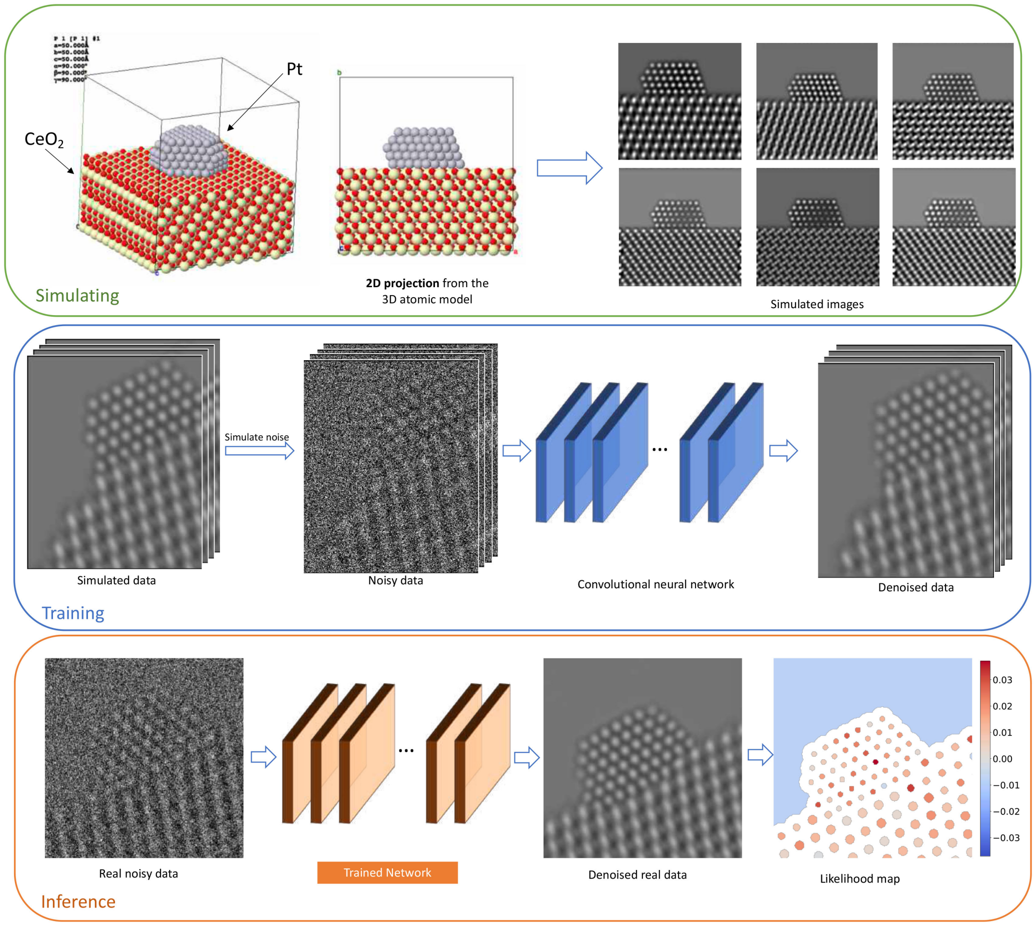

Fig. 2. Simulation-based denoising framework. (Top) A training dataset is generated by simulating TEM images of different structures at varying imaging conditions. Here we focus on structures of Pt nanoparticles supported on CeO_2. (Middle) A CNN is trained using the simulated images, paired with noisy 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 structures of interest, a likelihood map is generated to quantify the agreement between this structure and the noisy data.

small changes in the geometry of the image. Fig. SM7 shows that a denoising CNN can easily overfit the specific 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. 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.

B. Exploiting Non-Local Signal Structure

Images in scientific applications often have pixel-intensity distributions that differ significantly from those of photographic 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 photographic images have very small fields of view. For example, the field of view (or receptive field) [55] of DnCNN [5] and DURR [56] is 41×41 pixels, and 45×45 pixels respectively. Unlike photographic images, the TEM images in our case study exhibit very prominent global regularities, due to periodicity in the atomic structure of the imaged materials. In addition, electron-microscopy images are often measured at very low SNRs (in our case, the SNR for the real TEM data is about 3 dB, but most works for photographic images focus on an SNR above 22 dB, see e.g. [5]). As the SNR decreases, denoising CNNs tend to average over larger regions of the surrounding pixels, as demonstrated in [57] (qualitatively, this is the same behavior observed in a classical linear Wiener filter [58]). These considerations motivate using networks with large field of view to denoise TEM data.

Here we propose to denoise TEM data using deep networks with very large fields of view: 221×221 pixels and 893×893 pixels, a 25-fold and 400-fold increase with respect to generic denoising architectures respectively. In order to obtain a large field of view efficiently (i.e. without dramatically increasing the number of parameters in the network), we propose using a UNet network architecture [60]. We use 4 downsampling operations to achieve the 221×221 field of view and 6 downsampling operations to achieve the 893×893 field of view (see Section SM iv for a detailed description of the architecture). Table I compares the influence of the field of view in denoising photographic and our TEM images. For photographic images the performance of the network remains almost constant as we increase the field of view. In contrast, for TEM images increasing the field of

TABLE I

FIELD OF VIEW OF CNN ARCHITECTURES AND PERFORMANCE. MEAN PSNR AND SSIM ($\pm$ STANDARD DEVIATION) OF DIFFERENT CNN ARCHITECTURES ON THE (A) HELD-OUT SIMULATED TEST SET OF TEM DATA DESCRIBED IN SECTION V-B AND (B) VALIDATION SET OF THE DIV2K PHOTOGRAPHIC IMAGE DATASET [59]. FOR TEM IMAGES, INCREASING THE FIELD OF VIEW (FOV) 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. INCREASING THE NUMBER OF PARAMETERS WHILE KEEPING THE FIELD OF VIEW CONSTANT PRODUCES A MODEST GAIN IN PERFORMANCE. IN CONTRAST, CHANGING THE FIELD OF VIEW HAS ALMOST NO EFFECT ON DENOISING PHOTOGRAPHIC IMAGES. THE NETWORKS USED FOR PHOTOGRAPHIC IMAGES WERE TRAINED ON 512×512 PATCHES EXTRACTED FROM TRAINING IMAGES OF DIV2K [59] CORRUPTED WITH ADDITIVE GAUSSIAN NOISE WITH STANDARD DEVIATION $\sigma \in [0, 100]$

(a) TEM Images

MODEL	Parameters	Field of View	PSNR	SSIM
SBD + DnCNN [5]	668K	41×41	30.47 ± 0.64	0.93 ± 0.01
SBD + Small UNet [56]	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	38.05 ± 0.81	0.99 ± 0.01
SBD + UNet (128 base channels)	70.15M	893×893	42.87 ± 1.45	0.99 ± 0.01

(b) Photographic Images

MODEL	Parameters	Field of View	PSNR		SSIM	
			$\sigma = 30$	$\sigma = 70$	$\sigma = 30$	$\sigma = 70$
UNet	102K	49×49	29.67 ± 2.84	26.16 ± 2.79	0.83 ± 0.06	0.70 ± 0.09
UNet	352K	221×221	29.65 ± 2.76	26.08 ± 2.68	0.83 ± 0.05	0.70 ± 0.08
UNet	4.4M	893×893	29.54 ± 2.82	26.07 ± 2.80	0.83 ± 0.06	0.70 ± 0.09

view produces a dramatic improvement in performance (6 dB and 10 dB, when the field of view is 221×221 and 893×893 respectively). Increasing the number of parameters, while keeping the field of view constant, has a very modest effect, which suggests that the increase in field of view is the primary reason for the improvement.

In order to gain some insight into the denoising mechanisms learned by our models, we apply the gradient-based analysis proposed by [57]. We visualize the linear term in the first-order Taylor decomposition of the denoising map with respect to its input for specific pixels. In more detail, we compute the gradient of a pixel in the denoised image $f(y)_i$ with respect to the input noisy image y. This vector (or image) $\nabla_y(f(y)_i)$, makes it possible to visualize the influence of different regions of the noisy image on the denoised pixel $f(y)_i$. This approach is similar to visualization methods proposed in the context of image classification (e.g. [61], [62]). Our analysis reveals that the network learns to simultaneously exploit local structure as well as non-local periodicities in the data (see Fig. SM9). This demonstrates the remarkable flexibility of data-driven denoising based on deep learning.

C. 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 [63], or to the vacuum.
2) Fit a low-dimensional model to the denoised image within the region of interest. The low-dimensional model provides an estimate of the image value x_i at each pixel location $i \in \mathcal{R}$. In our case, we assume that pixel intensities within a given atomic column 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 independent and individually distributed (iid) Poisson (see Section SM iii), so the likelihood is given by

$$\mathcal{L}(\mathcal{R}) := \prod_{i \in \mathcal{R}} p_{x_i}(y_i), \tag{3}$$

where y_i denotes the noisy value in the ith pixel, and p_{x_i} is a Poisson probability mass function (pmf) with rate parameter x_i. Note that in the low-dimensional model, which assumes constant intensities, x_i is constant for all pixels in $\mathcal{R}$.

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 artefact 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 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

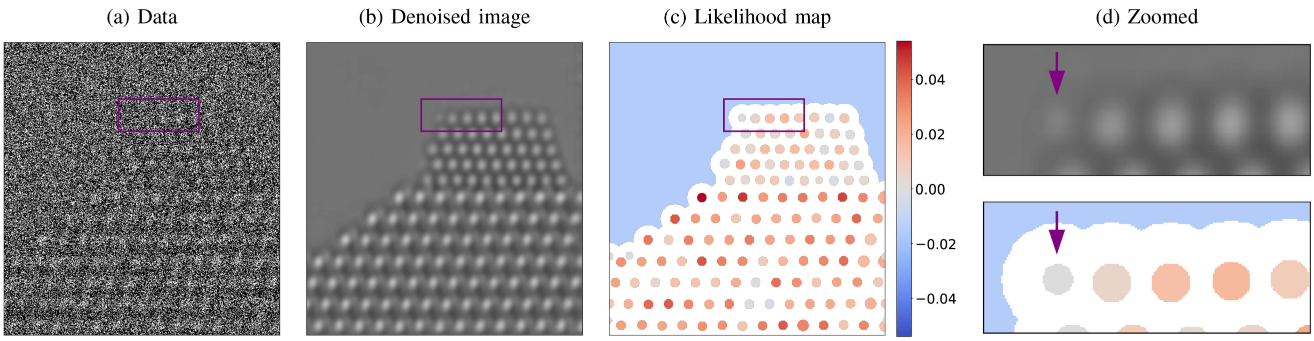

Fig. 3. Likelihood map. When the simulated noisy image in (a) is denoised using the proposed framework (b), a spurious atom appears at the left edge of the nanoparticle (see zoomed image (d)). The value of the likelihood map (c) at that location is very low, indicating that the presence of an atom is less consistent with the observed data than its absence.

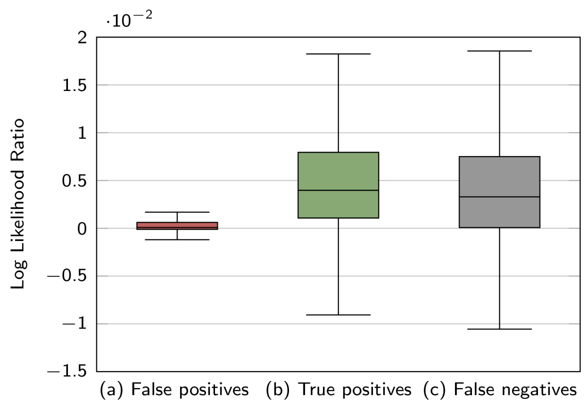

Fig. 4. Distribution of likelihood ratio. The figure shows the distribution of log-likelihood ratio of over 25,000 regions of interest computed from the surface of 1550 denoised images using the dataset described in Section V-C2. The empirical distribution is visualized as a box plot indicating the median, 25th quartile, 75th quartile, minimum and maximum value of the distribution. The regions containing spurious atoms (false positives, (a)) have a much lower log-likelihood ratio than the regions containing accurately recovered atoms (true positives, (b)). Regions where existing atoms were not detected (false negatives, (c)) have a higher log-likelihood ratio, comparable to that of the regions with accurately recovered atoms. The occurrence of missing and spurious atoms in denoised images is relatively low: out of the 25,732 regions of interest, only 2,457 and 2,368 were false positives and false negatives respectively.

interest. This is equivalent to performing a log-likelihood ratio test. We call this visualization a *likelihood map*.

Figs. 1 and 3 show likelihood maps for the real data and for a simulated example. In the simulated example (Fig. 3), a spurious atom is detected at the left end of the zoomed region. However, the likelihood map at that location is very low, which indicates that the presence of an atom may not be consistent with the observed data at that location. Fig. 4 shows the distribution of log-likelihood ratio of over 25,000 regions of interest extracted from the surface of over 1,550 denoised images obtained from the dataset described in Section V-C2. As shown in Fig. 4, the log-likelihood ratio values of spurious atoms (false postives, (a)) are much lower than those of correctly-identified atoms (true positives, (c)). When the network fails to detects atoms (false negatives, (b)), we observe that the log-likelihood ratio in such regions tends to be high. It is worth noting that the occurrence of spurious and missing atoms in the denoised images is relatively low: out of the 25,732 regions identified, only 2,457 and 2,368 regions correspond to spurious and missing atoms respectively.

Visualizing the likelihood is useful to quantify the agreement between the output of deep-learning models and the observed data, but it is important to note that the approach suffers from sampling bias. We focus on regions of the input that have been selected because they resemble structures of scientific interest. The data in those regions are therefore more likely to be in agreement with the presence of such structures, just by the sheer fact that they have been selected. This is a manifestation of the notorious multiple-comparisons problem [64], [65]. Overcoming this issue is an important challenge for future research.

IV. Dataset

The TEM image data used in this work correspond to images from a widely utilized catalytic system, which consist of platinum (Pt) nanoparticles supported on a larger cerium (IV) oxide (CeO_2) nanoparticle. 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 [66]–[68]. From a general point of view, this system can be considered as a model for supported nanoparticle catalysts, since a large number of heterogeneous catalysts are based on 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).

A. Real Data

The real data used to test the proposed SBD framework consist 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.

B. Simulation Dataset

The simulated TEM image dataset was generated using the multi-slice TEM image simulation method, as implemented in the Dr. Probe software package [69] (see Section SM i-A 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 s experimental frame (i.e., 0.45 counts per pixel in the vacuum region).

In the type of phase-contrast 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 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., Fig. SM1). 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 SM i-B and SM i-C.

V. Experiments and Results

In this section, we evaluate the performance of our proposed methodology and show that we outperform other methods by a large margin (more than 12 dB in PSNR on held-out simulated data). We also perform a thorough analysis of the generalization capability of our models, demonstrating that the CNNs are robust to variations in imaging parameters and in underlying signal structure. Furthermore, we demonstrate that standard performance metrics for photographs, such as peak signal-to-noise ratio (PSNR) and SSIM [18], may fail to produce a scientifically-meaningful evaluation of the denoising results, and we propose a few alternative metrics to remedy this. Finally, we show that our approach achieves effective denoising of real experimental data.

We use CNNs with the proposed UNet architecture with 128 base channels and 6 scales in all of our experiments (see Sections III-B and SM iv for more details). The networks were trained on 400×400 patches extracted from the training images and augmented with horizontal flipping, vertical flipping, random rotations between -45° and $+45^\circ$, and random resizing by a factor of 0.75-0.82. The models were trained using the Adam optimizer [70], 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. The details of training, validation and test data for each experiment are provided in the corresponding section. Since the models are trained on 400×400 patches, when applying them to larger images we divide the images into overlapping 400×400 patches, denoise them, and then combine them via averaging.

A. Generalization to Unseen Structures and Acquisition Conditions

In order to study the generalization ability of the proposed approach across different imaging parameters and signal structures we divided the simulated dataset described in Section IV 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 experts[2] (see Fig. SM1 in the supplementary material). The nanoparticle structure was classified into four categories, "PtNp1" through "PtNp4". PtNp1 and PtNp2 correspond to supported Pt nanoparticles of size 2 nm, which differ in the presence or absence 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. 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 SM i-A and in particular in Fig. SM5. D0 is the initial structure, D1/D2 a structure in which 1/2 atomic columns 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 performance of SBD is robust to variations in imaging parameters and in the underlying signal structure, as shown in Fig. SM8. We only observe a significant decrease in performance when the network is trained on black-contrast images and tested on other contrasts (interestingly the network generalizes well from white and intermediate contrasts to black contrasts).

B. Comparison of SBD With Other Methods

The imaging parameters of the real data, described in Section IV, are well described by the white contrast category defined in Section SM i-B. We therefore used the subset of simulated dataset corresponding to this contrast (5583 images) to compare

[2]Domain experts refers to three material scientists who specialize in TEM.

TABLE II
RESULTS ON SIMULATED TEST DATA. MEAN PSNR AND SSIM (± STANDARD DEVIATION) OF DIFFERENT DENOISING METHODS ON THE HELD-OUT SIMULATED TEST SET DESCRIBED IN SECTION V-B. SBD APPROACHES ACHIEVE THE BEST RESULTS. SBD COMBINED WITH THE PROPOSED ARCHITECTURE OUTPERFORMS ALL OTHER TECHNIQUES BY ABOUT 12 DB. THE PERFORMANCE OF SBD APPLIED TO ADDITIONAL ARCHITECTURES IS REPORTED IN TABLE I, AND DENOISED IMAGES ARE SHOWN IN FIG.S SM12 AND SM13

METHODS	PSNR	SSIM
Raw	3.56 ± 0.03	0.00 ± 0.00
Low Pass Filter [20]	21.59 ± 0.07	0.44 ± 0.03
Adaptive Wiener Filter [71]	22.42 ± 1.08	0.63 ± 0.02
VST + NLM [72]	26.55 ± 0.16	0.73 ± 0.01
VST + BM3D [73]	25.27 ± 0.15	0.80 ± 0.01
PURE-LET [7]	28.36 ± 0.88	0.93 ± 0.01
SBD + DnCNN [5]	30.47 ± 0.64	0.93 ± 0.01
SBD + Small UNet [56]	30.87 ± 0.56	0.93 ± 0.01
SBD + Proposed Architecture	**42.87** ± 1.45	**0.99** ± 0.01

our proposed methodology to other models. 90% of the data were used for training. The remaining 559 images were evenly split into validation and test sets. We compare our proposed UNet architecture (see Section SM iv) with two state-of-the-art architectures for photographic-image denoising [5], [56] (see Sections SM iv-B and SM iv-C), and with several classical denoising methods: low-pass filtering [20], adaptive Wiener filtering [71], BM3D [73], non-local means [7] and, a wavelet-based method known as PURE-LET [7]. A detailed description of these techniques is provided in Section SMv. For all methods, hyperparameters were chosen based on the validation data. Performance was measured in terms of SSIM [18] and peak signal-to-noise ratio (PSNR).

The results demonstrate that SBD is an effective denoising methodology for TEM data. Our proposed CNN outperforms all other methods by a margin of 12 dB in PSNR on the simulated test data, as shown in Table II, and Fig.s SM12 and SM13. SBD recovers the overall shape of the nanoparticle, the interface between the nanoparticle and the support, and the different periodic patterns of the CeO_2 support and Pt nanoparticle. Contrast features, such as subtle patterns of bright, intermediate and dark features associated with the atomic structure of the CeO_2 crystal, are well reproduced in the images denoised via SBD, but are mostly absent from the results of the baseline approaches.

C. Beyond PSNR: Towards Scientifically-Meaningful Evaluation Metrics

Domain scientists denoise images in order to extract scientifically relevant information. In our case, the atoms on the surface of nanoparticles are of particular interest, because the atomic configuration at the surface regulates the nanoparticle's ability to catalyze chemical reactions. It is therefore of critical importance to understand how different denoising methods recover these atoms. We can verify visually that SBD achieves a largely successful recovery in held-out simulated data, whereas the baseline methods described in Section V-B do not (see Fig. SM12 for example). However, visual inspection is a limited and non-quantitative evaluation tool. Unfortunately, standard metrics like PSNR and SSIM are insensitive to changes in the atomic structure of the nanoparticle surface, because these changes have a small effect on the overall intensity of the images. We demonstrate the lack of sensitivity through a synthetic example in Fig. SM14: when we add or remove an atom in the surface the PSNR and SSIM remain roughly constant. Motivated by the need for scientifically-relevant performance evaluation, we propose several metrics explicitly designed to account for changes in surface atomic configuration in Section V-C1. We report an evaluation of SBD using these metrics on a challenging test dataset in Section V-C2.

1) Evaluation Metrics: To define metrics that evaluate detection of surface atoms, we assume that there is a predefined approach to perform detection based on the denoised images. In our case of interest, we apply a blob detection algorithm (Laplacian of Gaussian [63]) to locate the centers, and compute the α-shape of all the atom centers using Delaunay triangulation [74]. Let A and B be the set of surface atoms of interest in the denoised image and the ground-truth clean image respectively. We propose the following four metrics to measure the fidelity of the recovered structure:

- **Precision** is the fraction of atoms in the denoised image that are also present in the clean image.

$$P(A, B) = \frac{|A \cap B|}{|B|}. \tag{4}$$

- **Recall** is the fraction of atoms in the clean image that are correctly recovered in the denoised image.

$$R(A, B) = \frac{|A \cap B|}{|A|}. \tag{5}$$

- **F1 score** combines precision and recall by giving them equal importance.

$$F(A, B) = 2\frac{P(A, B)R(A, B)}{P(A, B) + R(A, B)}. \tag{6}$$

- **Jaccard index** is an alternative measure consisting of the ratio between the size of the intersection between the recovered atoms and the ground truth divided by the size of their union.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{7}$$

When performing intersection and union operations, we consider two atoms to be the same if the distance between their centers is less than a threshold of 10 pixels. For comparison, the physical distance between neighboring atoms is about 0.16 nm, and 10 pixels correspond to a distance of 0.061 nm. All our metrics take values between 0 and 1 (1 is best). Fig. SM14 shows a synthetic example comparing three images with different atomic configurations: all the images have similar PSNR and SSIM values, but the precision, recall, F1 and Jaccard index show substantial differences.

2) Evaluating Atom Detection Accuracy: To evaluate the performance of the proposed approach to recover atoms at the surface, we designed a new dataset with 308 images, where the imaging parameters are set based on the real dataset described

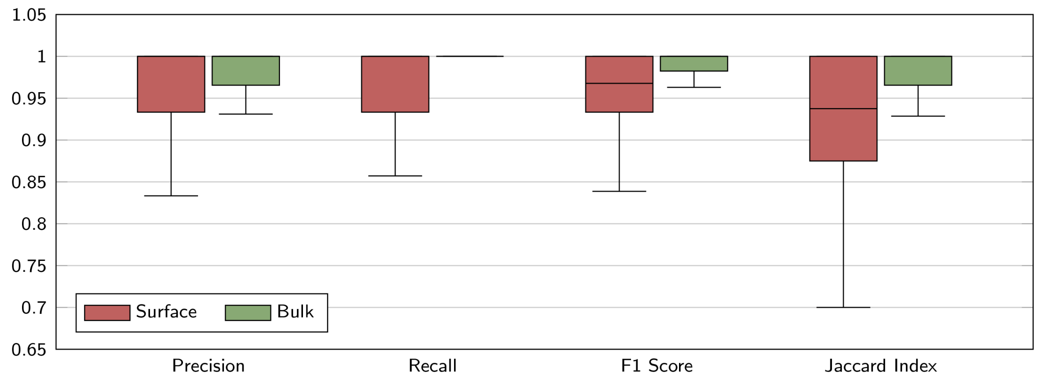

Fig. 5. Performance of SBD in terms of our proposed metrics. We compute all our proposed metrics (see Section V-C1) on over 7,000 denoised images corresponding to 25 unique noisy images sampled from the 308 clean images described in Section V-C2. The empirical distribution on the surface (red) and bulk (green) is visualized as box plots indicating the median, 25^{th} quartile, 75^{th} quartile, minimum and maximum value of the distribution. SBD has a near perfect performance in the bulk with all metric values hovering around 1. On the surface, SBD achieves a median score of 1 for precision and recall, and about 0.95 for F1 score and Jaccard index.

in Section IV-A. This new dataset is similar to the one used for the generalization experiments in Section V-A, but here we add more diverse surface defects. We created a series of 44 Pt/CeO_2 structural models with atomic-level surface defects such as the removal of an atom from a column, removal of two atoms, removal of all but one atom and addition of an atom at a new site (see Fig. SM15 for a visual overview). We hypothesize that these defects emulate dynamic atomic-level reconfigurations that could potentially be observed in real experiments. To match the image contrast of our real data, we simulated images under defocus values ranging from 6 nm to 10 nm, all with a tilt of 3° in x and -1° in y and a support thickness of 4 nm. SBD recovers all the atoms in the bulk almost perfectly, as reflected in the different metrics. On the surface, SBD achieves a median score of 1 for precision and recall, and more than 0.95 on F1 and Jaccard index (see Fig. 5).

D. *Performance on Real Data*

In the experiments reported in Sections V-B and V-C we used a network trained on all simulated images from the white contrast category defined in Section V-A. However, the real data described in Section IV more closely corresponds to a subset of white contrast images satisfying the following conditions: structure limited to PtNP2, thickness between 40 Å – 60 Å and, defocus between 5 nm and 10 nm. We used 236 images from this subset for training, and another such 15 images for validation. We also trained two state-of-the-art architectures for photographic image denoising - DnCNN [5] and DURR [56] on these data.

Results on real experimental data obtained using SBD trained on this relevant subset of white contrast are shown in Figs 1, SM16, and SM17. SBD produces denoised images that are of much higher quality than those of the baseline methods described in Section V-B, which contain obvious artefacts. Further, we validate the denoising results of SBD by comparing to an estimated reference image obtained by temporal averaging. Our real dataset consists of 40 frames that are approximately stationary and aligned. Therefore, their temporal average provides a good estimate for the ground-truth images. As shown in Fig. 6, the denoised intensity values of the atomic column approximately match those of the estimated reference image.

In the rest of this section, we compare the performance of SBD and unsupervised denoising techniques on the real experimental data, and analyze the effect of the design of the training dataset on the denoised output produced by SBD.

1) Comparison to Unsupervised Deep Denoising Methods: Unsupervised denoising techniques can be used to train a denoising CNN using only noisy images (see Section II for a discussion on this methodology). We apply the following unsupervised methods to the real data described in Section IV-A:

- **Noise2Noise** [39] is a strategy used to train CNN by using pairs of noisy images which correspond to the same clean image. We applied this method to our data by treating images captured in consecutive time steps as different noisy realizations of an underlying clean image. The results (shown in Fig. 7(b)) contain visible artefacts and missing atoms.
- **Blind-spot net** [43] is a CNN which is constrained to predict the intensity of a pixel as a function of the noisy pixels in its neighbourhood, without using the pixel itself. This method is competitive with the current supervised state-of-the-art CNN on photographic images. However, when applied to our real dataset it produces denoised images with visible artefacts (see Fig. 7(c)). A possible explanation is the limited amount of data (40 noisy images) we train on. To validate this hypothesis, we trained a blind-spot net on simulated training sets of different sizes. The performance on held-out data is indeed poor when the training set is small, but it improves to the level of supervised approaches as we use more training data (see Fig. SM18).
- Blind-spot net with early stopping. In Ref. [75] it is shown that early stopping based on noisy held-out data can boost the performance of blind-spot nets. Here we used 35 images for training the blind-spot net and the remaining 5 images

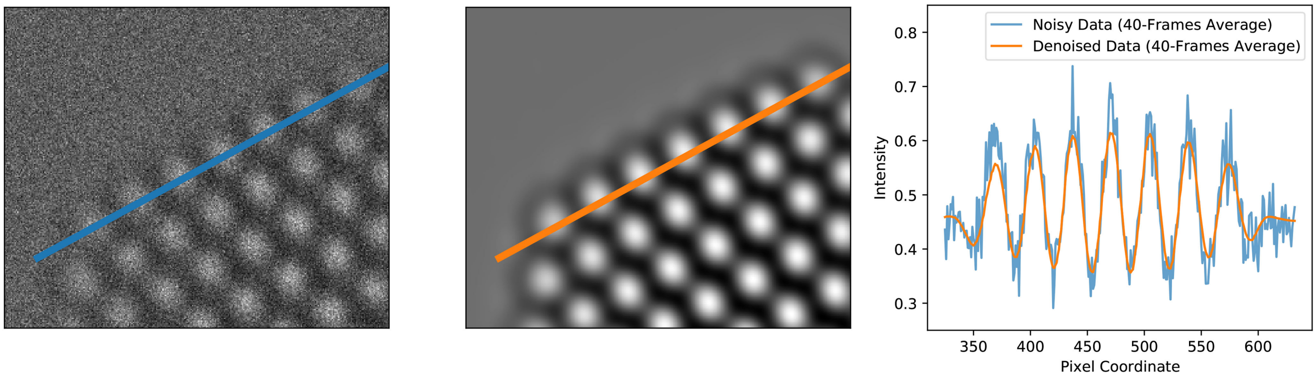

Fig. 6. Validation on real data. The real data consist of 40 frames which are approximately stationary and aligned. Their temporal average (left) therefore provides a reasonable estimate for the true intensity profile. In the image on the right, we compare the average intensity profile on the surface atomic columns of the platinum nanoparticle for the denoised data (middle) and the temporal average (left). The profiles are very similar (except for some spurious fluctuations in the temporal average), which suggests that the proposed approach achieves effective denoising on the real data.

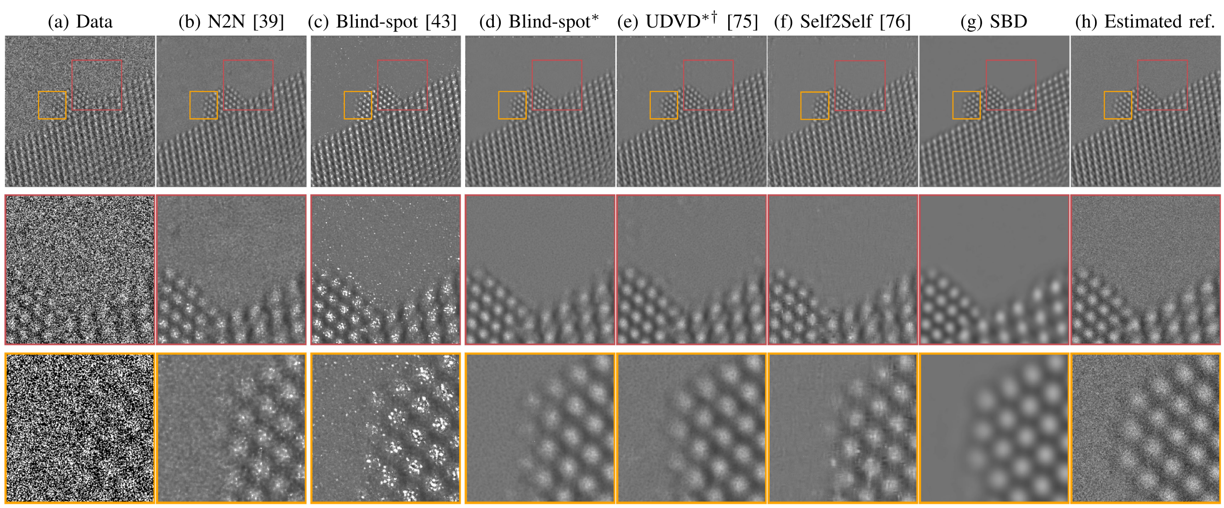

Fig. 7. Comparison of unsupervised denoising methods with SBD on real data. The real data described in Section IV-A denoised using SBD and unsupervised methods described in Section V-D1. The second and third rows zoom in on the region in red and green boxes respectively. Our proposed method denoises the real data more effectively than the unsupervised approaches. SBD is able to precisely recover the structure of the nanoparticle and has very few artefacts (compare visually to the estimated reference image obtained via time averaging; there are three missing atoms for most unsupervised methods in the third row). A $*$ indicates that the method used early stopping and † indicates that the method uses 5 noisy frames as input.

as a held-out validation set. We chose the model parameters that minimized the mean squared error between the noisy validation images and the corresponding denoised estimates. The results (shown in Fig. 7(d)) are significantly better than those of the standard blind-spot network. However, there are still noticeable artefacts, which include missing atoms.

- **Unsupervised Deep Video Denoising (UDVD)** [75] is an unsupervised method for denoising video data based on the blind-spot approach. It estimates a denoised frame using 5 consecutive noisy frames around it. Our real data consists of 40 frames acquired sequentially. UDVD produces better results than blind-spot net, but still contains visible artefacts, including missing atoms (see Fig. 7(e)). Note that, UDVD uses 5 noisy images as input, and thus has more context to perform denoising than the other methods (including SBD).
- **Self2Self** [76] is an unsupervised method specifically designed for denoising based on a single noisy image. This approach achieves near state-of-the-art performance in noisy photographic images corrupted with moderate amounts of noise [76]. However, when applied to our data, Self2Self produces images with clear artefacts; some of the atoms are missing and the shape of atoms are distorted (see Fig. 7(f)).

It is important to note that the backbone architectures of all these methods are UNets with large fields of view, like the one used for SBD. In our experiments, we trained the blind-spot nets and UDVD on 600×600 patches extracted from the real data. We used Adam optimizer [70] with a starting learning rate of 1×10^{-4} which was reduced in half for every 2000 epochs. We trained for a total of 5000 epochs. When performing early stopping, we picked the checkpoint with the best mean squared error on the validation set. Following [76], Self2Self was trained for 150,000 steps with the Adam optimizer and a starting learning rate of 10^{-4}.

As shown in Fig. 7, the unsupervised denoising methods produce higher-quality reconstructions than those of the baseline methods discussed in Section V-B (see Fig. SM16). However, they still suffer from visible artefacts, particularly on the surface of the nanoparticle, limiting their practical utility. UDVD is the method that achieves best performance, but it requires multiple noisy frames as input. In contrast, SBD can denoise the image effectively from a single noisy input frame (see Fig. 7(g)), as long as the simulated training data correspond closely to the real noisy image. Using a single frame is important in some applications, such as our case of interest, where the ultimate goal is to identify dynamic changes in the atomic structure of the nanoparticle.

2) A Word of Caution. Effect of Training Data on SBD: Figs SM16, SM17 and 7 show that SBD achieves impressive results on real data, but it is important to point out that this requires a careful design of the training dataset. Our real data broadly corresponds to images in the white contrast category, defined in Section V-A. However, when a network trained on white contrast images (Section V-B) is evaluated on the real data, it produces unnatural streak patterns in the bulk (see third row in Fig. SM19). When visually comparing this to the pattern in the bulk of the reference image computed by time averaging, it is evident that this is an artefact of denoising. This can be remedied by training the network on the more restricted subset of images described in Section V-D (see third row in Fig. SM19), whose imaging parameters are more suited to the real acquisition conditions. Since unsupervised denoising methods directly train on the real data, they do not suffer from this problem of mismatch between training and test data. The patterns recovered by unsupervised methods in the bulk are close to the estimated reference image (see Fig. SM19). However, as discussed in Section V-D1, they show significant artefacts on the surface of nanoparticle. Domain generalization, when there is a gap between the distribution of the training and test data is a fundamental challenge in machine learning [77]. Ref. [78] proposes a method to address this gap for denoising. They pre-train a CNN model on simulated data, and adjust a small set of CNN parameters adaptively and selectively for each individual experimental test image.

VI. Discussion and Conclusions

Our case study is a proof of concept that CNNs trained for denoising on simulated data can be remarkably effective when applied to real imaging data. It provides several insights and suggests future research directions that are relevant, beyond electron microscopy, to other domains where the images of interest can be simulated, such as medical imaging [29], [31], other types of microscopy [32], [33], or astronomy [79]. We show that the design of the training dataset is critical, so an important question is how to design simulated training 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. We also demonstrate that architectures tailored to photographic imaging can perform poorly when applied to other data. Designing CNNs for other domains requires an understanding of the image features that are exploited for denoising. Gradient visualization is shown to be useful here, but more advanced visualization techniques are needed. In addition, we demonstrate that standard metrics used to quantify performance in photographs may not be sensitive to scientifically relevant features, and propose several new metrics to address this problem. Although SBD outperforms other methods by a large margin, some artefacts 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.

Acknowledgment

All authors participated in writing and reviewing the manuscript. The authors acknowledge ASU Research Computing and NYU HPC for providing high performance computing resources and the John M. Cowley Center for High Resolution Electron Microscopy at Arizona State University.

References

[1] J.-M. Zuo and J. Spence, *Advanced Transmission Electron Microscopy: Imaging and Diffraction in Nanoscience*. New York, NY, USA: Springer, 2017.

[2] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature Methods*, vol. 2, no. 12, pp. 910–919, 2005.

[3] I. S. McLean, *Electronic Imaging in Astronomy: Detectors and Instrumentation*. Berlin/Heidelberg, Germany: Springer, 2008.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[6] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

[7] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed poisson–Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[8] D. Smith, "Characterization of nanomaterials using transmission electron microscopy," in *Nanocharacterisation*, vol. 37, Jan. 2015, ch. 1, pp. 1–29.

[9] F. Tao and P. Crozier, "Atomic-scale observations of catalyst structures under reaction conditions and during catalysis," *Chem. Rev.*, vol. 116, no. 6, pp. 3487–3539, Mar. 2016.

[10] X.-C. Bai, G. McMullan, and S. H. Scheres, "How CRYO-EM is revolutionizing structural biology," *Trends Biochem. Sci.*, vol. 40, no. 1, pp. 49–57, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S096800041400187X

[11] A. Faruqi and G. McMullan, "Direct imaging detectors for electron microscopy," *Nucl. Instrum. Methods Phys. Res. Sect. A: Accelerators, Spectrometers, Detectors Assoc. Equip., Radiat. Imag. Techn. Appl.*, vol. 878, pp. 180–190, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900217307787

[12] P. Ercius *et al.*, "The 4D camera—An 87 kHz frame-rate detector for counted 4D-STEM experiments," *Microsc. Microanal.*, vol. 26, no. S2, pp. 1896–1897, 2020.

[13] G. Sun, A. N. Alexandrova, and P. Sautet, "Structural rearrangements of subnanometer CU oxide clusters govern catalytic oxidation," *ACS Catalysis*, vol. 10, no. 9, pp. 5309–5317, 2020. [Online]. Available: https://doi.org/10.1021/acscatal.0c00824

[14] H. Guo, P. Sautet, and A. N. Alexandrova, "Reagent-triggered isomerization of fluxional cluster catalyst via dynamic coupling," *J. Phys. Chem. Lett.*, vol. 11, no. 8, pp. 3089–3094, 2020. [Online]. Available: https://doi.org/10.1021/acs.jpclett.0c00548

[15] E. L. Lawrence, B. D. Levin, B. K. Miller, and P. A. Crozier, "Approaches to exploring spatio-temporal surface dynamics in nanoparticles with *in situ* transmission electron microscopy," *Microsc. Microanal.*, vol. 26, no. 1, pp. 86–94, 2020.

[16] B. D. Levin, E. L. Lawrence, and P. A. Crozier, "Tracking the picoscale spatial motion of atomic columns during dynamic structural change," *Ultramicroscopy*, vol. 213, 2020, Art. no. 112978. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399119303122

[17] P. A. Crozier, E. L. Lawrence, J. L. Vincent, and B. D. Levin, "Dynamic restructuring during processing: Approaches to higher temporal resolution," *Microsc. Microanal.*, vol. 25, no. S2, pp. 1464–1465, 2019.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[19] J. L. Vincent *et al.*, "Developing and evaluating deep neural network-based denoising for nanoparticle TEM images with ultra-low signal-to-noise," *Microsc. Microanal.*, vol. 27, no. 6 pp. 1431–1447, 2021.

[20] P. Nellist and S. Pennycook, "Accurate structure determination from image reconstruction in ADF stem," *J. Microsc.*, vol. 190, no. 1/2, pp. 159–170, 1998.

[21] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 1998, vol. 98, pp. 839–846.

[22] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Jan. 2013.

[23] W. Jiang, M. L. Baker, Q. Wu, C. Bajaj, and W. Chiu, "Applications of a bilateral denoising filter in biological electron microscopy," *J. Struct. Biol.*, vol. 144, no. 1/2, pp. 114–122, 2003.

[24] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, Nov. 2003.

[26] H. J. Zhu, B. C. Han, and B. Qiu, "Survey of astronomical image processing methods," in *Proc. Int. Conf. Image Graph.*, 2015, pp. 420–429.

[27] W. Meiniel, J.-C. Olivo-Marin, and E. D. Angelini, "Denoising of microscopy images: A review of the state-of-the-art, and a new sparsity-based method," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3842–3856, Aug. 2018.

[28] S. Beckouche, J.-L. Starck, and J. Fadili, "Astronomical image denoising using dictionary learning," *Astron. Astrophys.*, vol. 556, 2013, Art. no. A132.

[29] B. Kim, M. Han, H. Shim, and J. Baek, "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images," *Med. Phys.*, vol. 46, no. 9, pp. 3906–3923, 2019.

[30] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 153–161, Mar. 2019.

[31] D. Minarik, O. Enqvist, and E. Trägårdh, "Denoising of scintillation camera images using a deep convolutional neural network: A Monte Carlo simulation approach," *J. Nucl. Med.*, vol. 61, no. 2, pp. 298–303, 2020.

[32] E. Giannatou, G. Papavieros, V. Constantoudis, H. Papageorgiou, and E. Gogolides, "Deep learning denoising of SEM images towards noise-reduced LER measurements," *Microelectron. Eng.*, vol. 216, 2019, Art. no. 111051.

[33] R. K. Vasudevan and S. Jesse, "Deep learning as a tool for image denoising and drift correction," *Microsc. Microanal.*, vol. 25, no. S2, pp. 190–191, 2019.

[34] J. M. Ede and R. Beanland, "Improving electron micrograph signal-to-noise with an atrous convolutional encoder-decoder," *Ultramicroscopy*, vol. 202, pp. 18–25, 2019.

[35] B. Manifold, E. Thomas, A. T. Francis, A. H. Hill, and D. Fu, "Denoising of stimulated raman scattering microscopy images via deep learning," *Biomed. Opt. Exp.*, vol. 10, no. 8, pp. 3860–3874, 2019.

[36] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, pp. 593–606, Mar. 2007.

[37] Y. Hel-Or and D. Shaked, "A discriminative approach for wavelet denoising," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 443–457, Apr. 2008.

[38] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.

[39] J. Lehtinen *et al.*, "Noise2Noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, in Proceedings of Machine Learning Research vol. 80, J. Dy and A. Krause, Eds. Jul. 10–15, 2018, pp. 2965–2974, https://proceedings.mlr.press/v80/lehtinen18a.html

[40] T.-O. Buchholz, M. Jordan, G. Pigino, and F. Jug, "CRYO-CARE: Content-aware image restoration for cryo-transmission electron microscopy data," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 502–506.

[41] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2129–2137.

[42] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 524–533.

[43] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 6970–6980.

[44] A. Krull, T. Vičar, M. Prakash, M. Lalit, and F. Jug, "Probabilistic noise2void: Unsupervised content-aware denoising," *Front. Comput. Sci.*, vol. 2, 2020, Art. no. 5.

[45] W. Khademi, S. Rao, C. Minnerath, G. Hagen, and J. Ventura, "Self-supervised Poisson-Gaussian denoising," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 2131–2139.

[46] M. Prakash, M. Lalit, P. Tomancak, A. Krul, and F. Jug, "Fully unsupervised probabilistic noise2void," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, 2020, pp. 154–158.

[47] Y. Zhang *et al.*, "A poisson-Gaussian denoising dataset with real fluorescence microscopy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11710–11718.

[48] J. M. Ede, "Deep learning in electron microscopy," *Mach. Learn. Sci. Technol.*, vol. 2, no. 1, 2021, Art. no. 011004.

[49] A. Suveer, A. Gupta, G. Kylberg, and I.-M. Sintorn, "Super-resolution reconstruction of transmission electron microscopy images using deep learning," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 548–551.

[50] J. P. Horwath, D. N. Zakharov, R. Megret, and E. A. Stach, "Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images," *npj Comput. Mater.*, vol. 6, no. 1, pp. 1–9, 2020.

[51] J. Madsen *et al.*, "A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images," *Adv. Theory Simulations*, vol. 1, no. 8, 2018, Art. no. 1800037.

[52] M. Ragone, V. Yurkiv, B. Song, A. Ramsubramanian, R. Shahbazian-Yassar, and F. Mashayek, "Atomic column heights detection in metallic nanoparticles using deep convolutional learning," *Comput. Mater. Sci.*, vol. 180, 2020, Art. no. 109722.

[53] M. Ziatdinov *et al.*, "Deep learning of atomically resolved scanning transmission electron microscopy images: Chemical identification and tracking local transformations," *ACS Nano*, vol. 11, no. 12, pp. 12742–12752, 2017.

[54] M. Ziatdinov *et al.*, "Building and exploring libraries of atomic defects in graphene: Scanning transmission electron and scanning tunneling microscopy study," *Sci. Adv.*, vol. 5, no. 9, 2019, Art. no. eaaw 8989.

[55] A. Araujo, W. Norris, and J. Sim, "Computing receptive fields of convolutional neural networks," *Distill*, vol. 4, no. 11, 2019, Art. no. e21. [Online]. Available: https://distill.pub/2019/computing-receptive-fields

[56] X. Zhang, Y. Lu, J. Liu, and B. Dong, "Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration," in *Proc. 7th Int. Conf. Learn. Representations*, New Orleans, LA, USA, May 6–9, 2019, https://openreview.net/forum?id=SJfZKiC5FX

[57] S. Mohan, Z. Kadkhodaie, E. P. Simoncelli, and C. Fernandez-Granda, "Robust and interpretable blind image denoising via bias-free convolutional neural networks," in *Proc. Int. Conf. Learn. Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=HJlSmC4FPS

[58] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Eng. Appl*. Cambridge, MA, USA: MIT Press, 1950.

[59] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 126–135.

[60] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[61] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Workshop Int. Conf. Learn. Representations*, 2014.

[62] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Rec.*, vol. 65, pp. 211–222, 2017.

[63] T. Lindeberg, "Scale selection properties of generalized scale-space interest point detectors," *J. Math. Imag. Vis.*, vol. 46, no. 2, pp. 177–210, 2013.

[64] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. Ser. B. (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995. [Online]. Available: http://dx.doi.org/10.2307/2346101

[65] Y. Benjamini, "Simultaneous and selective inference: Current successes and future challenges," *Biometrical J.*, vol. 52, no. 6, pp. 708–721, 2010.

[66] T. Montini, M. Melchionna, M. Monai, and P. Fornasiero, "Fundamentals and catalytic applications of CeO_2-based materials," *Chem. Rev.*, vol. 116, no. 10, pp. 5987–6041, 2016.

[67] W. Yu, M. D. Porosoff, and J. G. Chen, "Review of Pt-based bimetallic catalysis: From model surfaces to supported catalysts," *Chem. Rev.*, vol. 112, no. 11, pp. 5780–5817, 2012.

[68] Y. Nie, L. Li, and Z. Wei, "Recent advancements in Pt and Pt-free catalysts for oxygen reduction reaction," *Chem. Soc. Rev.*, vol. 44, no. 8, pp. 2168–2201, 2015.

[69] J. Barthel, "Dr probe: A software for high-resolution stem image simulation," *Ultramicroscopy*, vol. 193, pp. 1–11, 2018.

[70] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 7–9, 2015, http://arxiv.org/abs/1412.6980

[71] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1990.

[72] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2005, vol. 2, pp. 60–65.

[73] M. Makitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91–103, Jan. 2013.

[74] F. P. Preparata and M. I. Shamos, "Convex hulls: Basic algorithms," in *Computational Geometry*. Berlin/Heidelberg, Germany: Springer, 1985, pp. 95–149.

[75] D. Y. Sheth *et al.*, "Unsupervised deep video denoising," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1759–1768.

[76] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1890–1898.

[77] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," in *Proc. 9th Int. Conf. Learn. Representations*, Virtual Event, Austria, May 3–7, 2021, https://openreview.net/forum?id=lQdXeXDoWtI

[78] S. Mohan, J. Vincent, R. Manzorro, P. Crozier, C. Fernandez-Granda, and E. Simoncelli, "Adaptive denoising via gaintuning," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 23727–23740, 2021.

[79] J. Peterson *et al.*, "Simulation of astronomical images from optical survey telescopes using a comprehensive photon Monte Carlo approach," *Astrophys. J. Suppl. Ser.*, vol. 218, no. 1, 2015, Art. no. 14.