ARTICLE   OPEN

Check for updates

Deep convolutional neural networks to restore single-shot electron microscopy images

I. Lobato[1,2 ✉], T. Friedrich [iD][1,2] and S. Van Aert [iD][1,2 ✉]

Advanced electron microscopy techniques, including scanning electron microscopes (SEM), scanning transmission electron microscopes (STEM), and transmission electron microscopes (TEM), have revolutionized imaging capabilities. However, achieving high-quality experimental images remains a challenge due to various distortions stemming from the instrumentation and external factors. These distortions, introduced at different stages of imaging, hinder the extraction of reliable quantitative insights. In this paper, we will discuss the main sources of distortion in TEM and S(T)EM images, develop models to describe them, and propose a method to correct these distortions using a convolutional neural network. We validate the effectiveness of our method on a range of simulated and experimental images, demonstrating its ability to significantly enhance the signal-to-noise ratio. This improvement leads to a more reliable extraction of quantitative structural information from the images. In summary, our findings offer a robust framework to enhance the quality of electron microscopy images, which in turn supports progress in structural analysis and quantification in materials science and biology.

INTRODUCTION

The quality of modern electron microscopes, such as SEM, STEM, and TEM, has greatly improved. However, the quality of the experimental images produced by these instruments is often compromised by stochastic and deterministic distortions arising from the instrument or its environment[1–3]. These distortions can occur during the acquisition, transmission, or reproduction of the image. Despite technical improvements in the design of high-performance electron microscopes[1–4], the presence of these distortions in the recorded images may hinder the extraction of quantitative information from the samples under study[5].

In TEM, images are acquired in a single shot using parallel acquisition. Here, the main sources of distortions are the detector noise, which is a combination of shot noise associated with the uncertainty of photon/electron detection, dark-current noise resulting from statistical variation in the number of thermally generated electrons within the detector, and readout noise resulting from the electronics that amplifies and digitizes the charge signal. Other sources of distortions for TEM include X-ray noise, which is produced by X-rays that saturate one or more nearby pixels as they pass through the detector[6,7], and dead pixel noise, which is caused by permanently damaged pixels on the sensor and often appears as black spots in the recorded images.

In S(T)EM, images are formed pixel by pixel by scanning a convergent electron beam across the sample and detecting the scattered, back-scattered or secondary electrons at each point. The main sources of distortions are the detector noise, which is a combination of shot noise hitting the scintillator, Gaussian noise resulting from the photomultiplier tube (PMT)[8], and readout noise from the electronics that amplifies and digitizes the electron signals. Unlike TEM imaging, the serial nature of SEM and STEM can introduce additional distortions into the resulting images due to time delays between measurements. At high doses, the main source of nonlinear distortion is the probe's fly-back time, where data collection pauses until scanning on the next line resumes.

This produces a net two-dimensional random displacement of the pixel row known as horizontal and vertical scan distortion. These nonlinear distortions can often be corrected using iterative algorithms that require a series of images[9,10] or a single image with a high-resolution periodic structure[11,12]. Moreover, S(T)EM images obtained through high-speed scans (dwell time < 1 μs[13]) may display a non-uniform scan speed along individual scan lines resulting in a smearing effect that produces another type of nonlinear distortion. While these distortions can be partly compensated for periodic structures[13], they cannot be fully compensated for arbitrary specimens. Other types of distortion include row-line noise, which is caused by the detector's non-response over a few pixels, and X-ray noise, which is produced by X-rays hitting the detector. These distortions can reduce the signal-to-noise ratio (SNR) and limit the amount of retrievable information about the electron-specimen interaction. Although increasing the beam current or acquisition time can improve the SNR, it can also increase other types of distortion, such as drift, translation, shear, rotation, expansion, or contraction of the entire image. Moreover, it is unsuitable for beam-sensitive materials or for dynamic imaging requiring a short exposure time for each frame. Lowering the electron dose can also decrease the quality of the recorded images and limit the reliability of structural information extracted from them.

Various algorithms have been developed to improve the SNR of electron microscopy (EM) images, including spatial filters such as median filters, Gaussian filters, Bragg filters, and Wiener filters[14–16]. More complex methods for denoising EM images include nonlinear iterative Wiener filtering algorithms[17] and block matching[18,19], although they can be computationally intensive. Another option for improving the SNR is to average a series of registered frames, using either rigid[20] or non-rigid[9,10] registration methods. However, these methods require a high overall electron dose and repeated recordings of the material. In addition, EM images often exhibit a combination of different types of

[1]Department of Physics, EMAT, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. [2]NANOlab Center of Excellence, Department of Physics, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. ✉email: Ivan.Lobato@uantwerpen.be; Sandra.VanAert@uantwerpen.be

npj

distortions due to several factors including the instrument environment, scan instabilities, scan speed, and dose. Therefore, there is a need for image restoration algorithms specifically designed for single-shot EM images.

In recent years, machine learning methods based on artificial neural networks, particularly convolutional neural networks (CNNs), have become the state-of-the-art approach for various tasks such as image classification[21], image segmentation[22], image denoising[23], image restoration[24], image deconvolution[25], and image super-resolution[26]. These methods, which involve adjusting the weight connections between neurons during training, have been made possible by the development of techniques such as the Rectified Linear Unit (ReLU)[27], dropout regularization[28], batch normalization[29], and improvements in GPU technology. While CNN-based approaches have achieved strong performance in denoising specific types of EM images, they are limited by their reliance on small simulated or experimental datasets and incomplete modeling of the various types of noise present in experimental EM data[30–35]. To the best of our knowledge, there is currently no algorithm that can effectively compensate for all types of distortion in a single-shot EM image without requiring retraining and regardless of the sample being studied.

In this study, we use a machine learning approach to restore EM images using a Concatenated Grouped Residual Dense Network (CGRDN) and a combination of loss functions and a generative adversarial network (GAN)[36]. This approach not only learns an end-to-end mapping between distorted and undistorted EM images, but also a loss function to train this mapping. Since we only have access to distorted data experimentally, we generate undistorted and distorted EM images by applying all distortions that can be corrected on single-shot EM images. By training the neural network to produce an undistorted output regardless of the level and combination of distortions in the input, it implicitly learns to detect and repair the distortions. This approach demonstrates impressive results for restoring both periodic and non-periodic specimens with different combinations of severe distortions. Importantly, the results show that both peak positions and intensities in atomic-resolution images can be reliably determined. In addition, the restoration time is only of the order of seconds for a 2k × 2k image.

RESULTS AND DISCUSSION

Electron microscopy techniques, such as SEM, STEM, and TEM, each possess distinct sources of noise and feature variations across both low and high resolutions. Consequently, we have trained our neural network architecture on an assortment of six diverse datasets, encompassing low-resolution (LR) and high-resolution (HR) images for each microscopy modality. Our empirical findings demonstrate that optimal performance is achieved by training separate neural networks for LR and HR features, particularly under low-dose conditions. In these scenarios, the neural network can capitalize on the specific feature distribution learned during its training phase. Moreover, our network achieves its best performance when processing raw, unmodified data, leveraging the prior knowledge acquired during training regarding the physics of electron detection for each microscopy modality. It merits emphasis that throughout the training and validation stages, we exclusively employed synthetic pairs of undistorted and distorted EM images. Nevertheless, our testing phase incorporated a heterogeneous set of images. For the atomic structure quantification process, we utilized the multislice method to generate the undistorted images, while employing our specialized noise model to produce their distorted counterparts. For other types of studies, we relied on experimental images. While our results primarily center on HR-STEM image restoration, this emphasis is due to the widespread use of HR-STEM for analyzing and quantifying atomic structures. However, our

findings are similarly transferable to other microscopy modalities. It is imperative to underscore that although a fully convolutional neural network is trained on a specific image size, it exhibits remarkable adaptability in handling images of varying dimensions. This versatility remains valid as long as the features identified during the training phase are manifested in the test image, irrespective of its size.

Implementation details and training

Moving on to implementation and training, the optimization of the neural network efficiency relies on selecting appropriate hyperparameter values. Therefore, we conducted an ablation study concerning model size, elaborated in the Supplementary Information. Specific network components, such as residual dense blocks (RDB) and grouped residual dense blocks (GRDB), are maintained at their maximum allowable values, which are constrained by our hardware capabilities. We set 4×4 filters for up/down-convolutional layers with strides of 2 and 1×1 filters for feature fusions. Otherwise, we used 3×3 filters with stride 1. Zero-padding was used for all convolutional layers. Each RDB has a growth rate of 32 channels, 8 pairs of convolutional layers, 4 RDB, 4 GRDB and ReLU activation. The initial convolution layer output has 64 channels and the final convolutional layer has 1 output channel. We would like to point out that the initial convolution layer output, growth rate and the number of convolutional layers in each RDB was obtained as trade-off between the accuracy and our hardware constraints. The above parameters produce a model of 152 layers with 7.04M parameters. The model is implemented using the Tensorflow 2.10 framework and is trained with a NVIDIA GTX Titan Volta GPU.

Our learning policy is based on Adam optimizer[37] with $\beta_1 = 0.5$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-7}$ and it is divided in three stages. In the non-adversarial first part, the generator is trained to minimize the $\mathscr{L}_1$ loss for 2 epochs with a learning rate of 5×10^{-5}. In order to prevent training instability and bad local optima, the learning rates were warmup for 65×10^3 steps[38,39]. This is followed by a second non-adversarial part in which the generator is trained to minimize the pixel-wise loss (Eq. (14)) with $\lambda_1 = 2.0$, $\lambda_2 = 1.0$, $\lambda_{mlwt} = 0.25$, $\lambda_{fs} = 0.10$, $\gamma = 0.125$, $\lambda_{mean} = 1.0$, and $\lambda_{std} = 0.5$. At this stage, the network was trained for 2 more epochs with a learning rate of 2.5×10^{-5}. In the third part, our generator is trained in adversarial manner as part of the cGAN to produce realistic images with $\lambda_{Adv} = 2 \times 10^{-5}$. We use a learning rate of 1.25×10^{-5} and 1×10^{-4} for the generator and discriminator, respectively, and halved at each epoch. For the adversarial training, we follow the standard approach from ref. [40] in which we alternate between one gradient descent step on D and G networks. In this part, the network was trained for 3 epochs. Furthermore, each epoch of the training process required three days of processing time on a single NVIDIA GTX Titan Volta GPU.

The mini-batch size is set to 4. Each of our six datasets included 2.5 million training examples and 100,000 validation examples, all of which were produced through our data generation process. The simulation of each image pair takes approximately 1.5 seconds, utilizing a single processing core operating at a speed of 3.5 GHz. The input data has a spatial size of 256 by 256 pixels. We encountered some issues during training, including loss spikes and instability, which were caused by the diversity of samples, orientations, distortions, and the combination of loss functions. To solve this issue, we implemented gradient clipping, which restricts the range of the gradients between –1 and 1, to prevent sudden spikes in loss. We would like to emphasize that throughout the training, we diligently monitored for signs of overfitting and found no instances, underscoring the reliability and robustness of our model. Details about the ablation study for the CGRDN architecture, which is based on the L1 metric as a function of the model size, are provided in the Supplementary Information.

Comparison to state-of-the-art algorithms

In this section, we compare the performance of three different image-denoising architectures: the Grouped Residual Dense Network (GRDN)[23], the Multi-resolution U-Net (MR-UNET)[31], and our proposed architecture, CGRDN. To evaluate the performance of these architectures, we use the widely recognized peak signal-to-noise ratio (PSNR) metric on the validation dataset, which is defined as follows:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right), \qquad (1)$$

where *MAX* denotes the maximum possible pixel value of the images, and *MSE* represents the mean squared error between the distorted and undistorted images. However, it is important to note that PSNR only measures the pixel-wise differences between the original and reconstructed images and does not account for other crucial factors such as visual perception and structural similarity. The GRDN architecture was previously ranked first in terms of PSNR and structure similarity index in the NTIRE2019 Image Denoising Challenge. The MR-UNET extends the functionality of the decoder in a U-Net[41] by adding additional convolutional layers to the hidden layers in order to produce coarse outputs that match low-frequency components. The results of our comparison are summarized in Table 1, which shows the number of parameters and the resulting PSNR for each architecture, and show that the GRDN and CGRDN are more efficient architectures because they require approximately seven times fewer parameters than the MR-UNET, while still

achieving a higher PSNR. It is interesting to note that our CGRDN architecture achieved a higher PSNR than the GRDN, while only requiring an additional 20,000 parameters.

We also compared the performance of our image restoration network to the Block-matching and 3D filtering (BM3D)[18] algorithm in terms of PSNR. BM3D is a widely used technique for removing noise from images through a process called denoising. It segments the image into overlapping blocks and identifies similar patterns among them to estimate the original image and reduce noise. BM3D has demonstrated effectiveness in denoising images with high levels of noise and serves as a benchmark for image-denoising algorithms in image processing. The average PSNR of BM3D and our network on the validation dataset was 30.45 and 36.96 dB, respectively. These results demonstrate that our network outperforms BM3D by a significant margin of 6.51 dB. Figure 1 illustrates the performance of our network and BM3D on two randomly generated, high-resolution STEM images with standard experimental noise values. These images were simulated using the procedure outlined in the Methods section. The figure displays the original distorted images Fig. 1a, e, and undistorted images Fig. 1d, h, as well as the denoised output from BM3D Fig. 1b, f, and the restored output from our network Fig. 1c, g.

These results demonstrate that our image restoration network significantly enhances image quality, as measured by PSNR. However, it is noteworthy that PSNR is not always a reliable indicator of image quality since it merely measures pixel-wise differences between original and reconstructed images and overlooks other critical factors such as visual perception and structural similarity. Hence, it is crucial to employ various image quality metrics, along with PSNR, to obtain a more comprehensive evaluation of the performance of image restoration techniques.

Atomic structure quantification

While the CNN was trained to restore images of a wide variety of imaging modes, STEM is of particular interest since it is routinely

Table 1. PSNR denoising performance comparison of different network architectures.

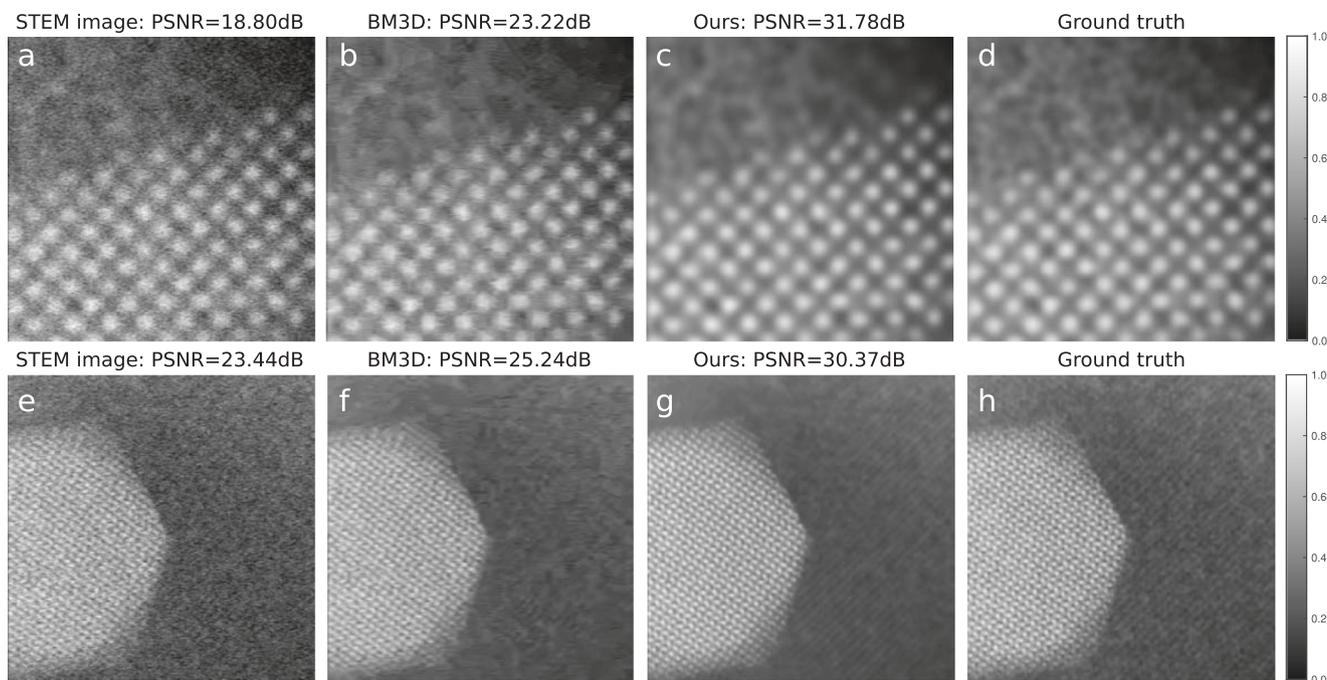| Method	# parameters	PSNR
MR-UNET[31]	51.7M	36.70 dB
GRDN[23]	7.02M	36.90 dB
CGRDN this work	7.04M	36.96 dB

Fig. 1 Comparison of the CNN restoration to the BM3D algorithm. CNN restoration results compared with BM3D in terms of PSNR for two random simulated STEM specimens using standard experimental noise values. **a, e** Simulated STEM images. **b, f** Output from BM3D. **c, g** Output from CNN. **d, h** Ground truth STEM images.
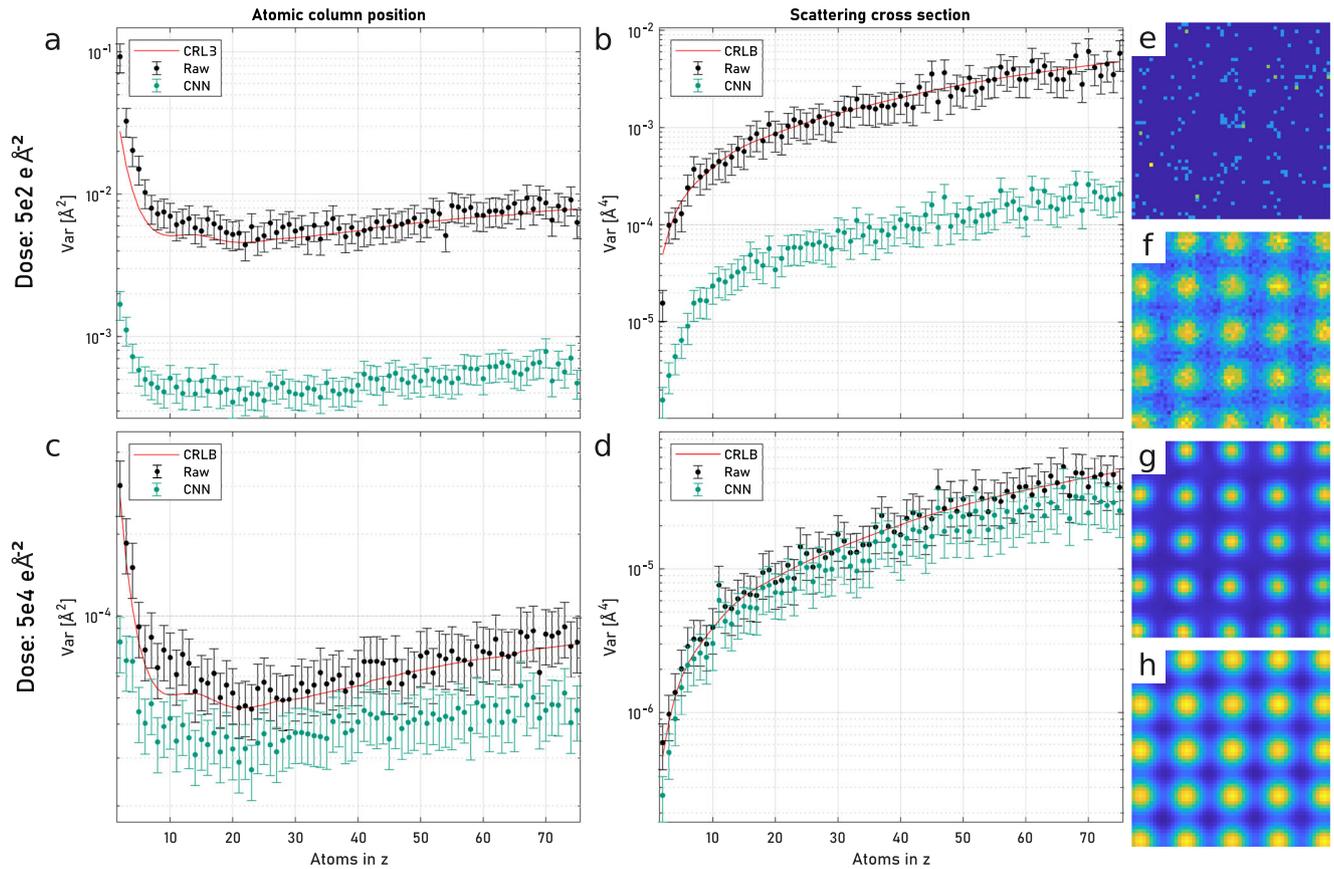
Fig. 2 Comparison of the precision in estimating unknown structure parameters with and without the use of the CNN. The precision of atomic column position and SCS measurements have been obtained from a series of Pt-bulk samples with a thickness varying from 2 to 75 atoms. **a** Precision of the atomic column locations for a dose of $5e^2e\text{Å}^{-2}$. **b** Precision of SCS measurements for a dose of $5e^2e\text{Å}^{-2}$. **c** Precision of atomic column locations for a dose of $5e^4\text{Å}^{-2}$. **d** Precision of SCS measurements for a dose of $5e^4\text{Å}^{-2}$. **e** Example of a raw STEM image at $z = 2$ and dose $= 5e^2e\text{Å}^{-2}$. **f** Example of a restored STEM image at $z = 2$ and dose $= 5e^2e\text{Å}^{-2}$. **g** Example of a raw STEM image at $z = 75$ and dose $= 5e^4e\text{Å}^{-2}$. **h** Example of a restored STEM image at $z = 75$ and dose $= 5e^4e\text{Å}^{-2}$. The error bars indicate 95% confidence intervals.

used for the quantification of atomic structures[42–44] in terms of atomic column positions and their corresponding scattering cross sections (SCS), which allows us to study the impact of the proposed image restoration method quantitatively. The probe position integrated scattering cross-section, short SCS, in atomic-resolution STEM images is defined as the integrated intensity of an atomic column, which is typically modeled as a 2D Gaussian function. Since the SCS scales with the atomic number $\approx Z^{1.7}$[45,46] and mostly increases monotonically with thickness for large collection angles, it is routinely used for atom counting. The evaluation of the effect of image restoration on the quantitative assessments of STEM images is done for three complementary case studies, using MULTEM[47,48] to create multislice simulations and the StatSTEM software for all model fittings[44]. All evaluations are based on 100 distortion/noise realizations for each dose setting.

1. We demonstrate the effect of image denoising with an idealized setup in analogy to the study conducted in ref. [44], where the precision of the determination of the location and SCS of an atomic column was determined over a wide range of signal-to-noise-ratios (SNRs) using pure Shot noise. This setting allows the comparison to the theoretical limit of variance for unbiased estimators, the so-called Cramér–Rao–Lower Bound (CRLB). The simulated STEM dataset is a bulk Pt crystal in [001] orientation and contains STEM images over 75 depth sections with unit cell spacing in z-direction.

2. A more practical example, that includes crystal irregularities, is chosen to determine the impact of a combination of noise, scan-line-distortions and fast-scan distortion. In this case, we evaluate the mean absolute error (MAE) for atomic column positions and the mean absolute percentage error (MPE) for the SCSs of atomic columns, as well as the variance of these measurements. This serves to show in particular the independence of the approach on the structural periodicity for atomic-resolution STEM images.

3. For a simulated Pt nanoparticle, it is demonstrated that distortion correction yields not only a more accurate localization of atomic columns but also enables more reliable atom counting.

Details regarding the simulation settings for all samples are provided in the Supplementary Information. The results of the first study are shown in Fig. 2. Examples of the underlying STEM images are given for the extremes of SNRs (i.e., the smallest thickness and lowest dose, and largest thickness and highest dose) for raw and restored images in Fig. 2e–h. Comparing Fig. 2e, f, it can be visually observed that even at a very low dose, the CNN can faithfully recover the underlying structure. The notable outcomes observed in these figures are primarily attributed to our extensive training dataset. This dataset enables the CNN to accurately model the Poisson noise distribution, inherently linked to the signal, providing the CNN with a significant advantage in terms of prior knowledge. It is essential to emphasize that our demonstration primarily considers Poisson noise as the sole source of distortion. In real-
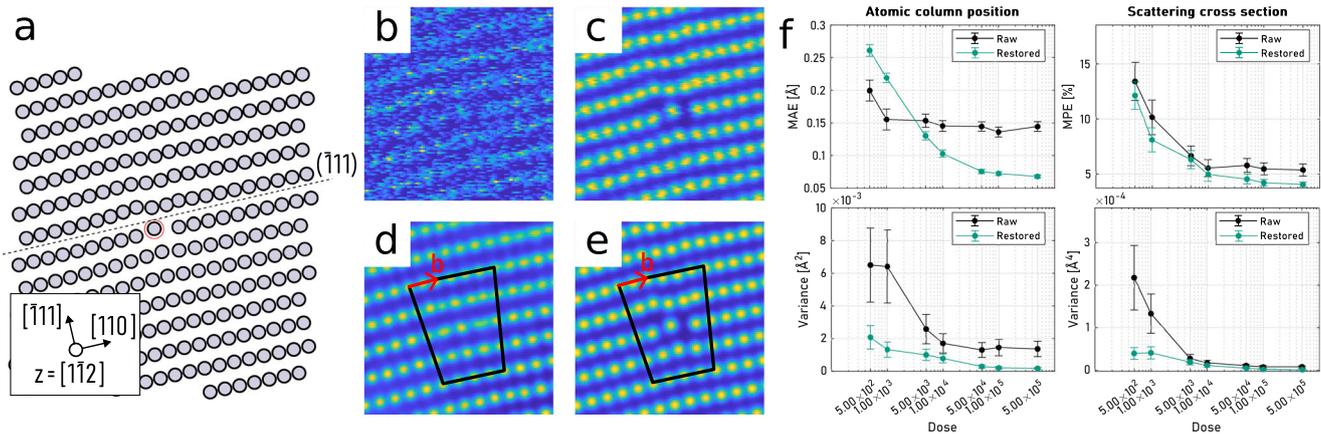
Fig. 3 Application of the CNN restoration to an imperfect crystal structure. a Schematic of the Pt structure in [112] zone axis with a unit edge dislocation of Burgers vector $b = 1/2[110]$ in the (111) glide plane. **b** Corrupted raw HAADF-STEM image with a dose of $5e^2e\text{Å}^{-2}$. **c** Corrupted raw image with a dose of $5e^5e\text{Å}^{-2}$. **d** Restored image with a dose of $5e^2e\text{Å}^{-2}$. **e** Restored image with a dose of $5e^5e\text{Å}^{-2}$. **f** Quantification results for the atomic column positions and scattering cross sections of the atomic columns around the center of the edge dislocation (marked with red circles in panel (**a**)). The error bars indicate 95% confidence intervals.

world scenarios, microscopy data often encounter various noise sources, which makes our demonstration somewhat idealized. The effect of the CNN is measurable both in terms of the precision with which atomic columns can be located, and in SCS measurement precision, and is particularly pronounced in the low dose range as illustrated in Fig. 2a, b. As the dose increases, the precision of the structural measurements of raw and restored data eventually converges (Fig. 2c, d). Interestingly, these results demonstrate that the theoretical precision limit, as defined by the CRLB, can be surpassed through the application of image restoration. It is vital to emphasize that such an observed outperformance should not automatically be interpreted as a cause for concern regarding overtraining. From a physics perspective, this outcome can be expected. In fact, we calculated the CRLB for the conventional case where no prior knowledge of the parameters is assumed. In contrast, it can be expected that prior knowledge is included in the CNN due to the extensive training it undergoes.

The restoration results in the first example arguably benefit from the underlying perfect crystal symmetry, which is why we also test the CNN for imperfect structures. The Pt-bulk model depicted in Fig. 3a is in [112] zone axis orientation, six unit cells thick and contains a unit edge dislocation of Burgers vector $b = 1/2[110]$ in the (111) glide plane; a dislocation commonly observed in fcc metals[49]. The structure was created using the Atomsk software, which determines atom positions corresponding to the displacement fields predicted by the elastic theory of dislocations[50]. The simulated HAADF-STEM images were subjected to varying noise levels from $5e^2e\text{Å}^{-2}$ to $5e^4e\text{Å}^{-2}$, and further corrupted by scan-line distortions as outlined in the Methods section. Example reconstructions for raw images at doses of $5e^2e\text{Å}^{-2}$ and $5e^4e\text{Å}^{-2}$ are shown in Fig. 3b, c, respectively, in Fig. 3d, e. In the low-dose raw image, individual atomic columns are hardly recognizable. Without the prior knowledge of the atomic column positions, any attempt at model fitting would have to overcome the challenge of performing reliable peak finding first, which is a factor not considered here. The reconstruction of this image, shown in Fig. 3d, on the other hand, displays very clear peaks. A Burgers circuit is superimposed on the image to highlight that despite the poor separation of columns in the raw image, the dislocation with its correct Burgers vector b is maintained, meaning that the structure as a whole is retrieved correctly, albeit the individual column positions may not be fully accurate as can be seen in the mean absolute position error of the columns around the center of the dislocation (columns within the red circle

in Fig. 3a for low doses shown in Fig. 3f. However, the error drops rapidly with increasing dose and shows a clear improvement against raw images. The position accuracy is therefore not only a result of denoising but also of the accurate correction of scan-line and fast-scan distortions. The comparatively high accuracy for the raw image fitting at low doses can be attributed to the fact that correct initial column positions are given for the fitting procedure. Since the column can hardly be located in the noisy images, the fitting algorithm on average does not move the position much away from this initial position. The CNN, on the other hand, reconstructs a clearly visible atomic column, but the available information in the underlying image is insufficient for accurate positioning. However, the proper retrieval of the dislocated atomic column at higher doses shows that the CNN is not by default just picking up on periodicity but faithfully recovers the atomic structure also in the presence of non-periodic features in atomic-resolution STEM images.

Also the SCS measurements improve in accuracy by the restoration, which would translate directly into improvements for atom counting studies. An example of such an atom counting scenario is presented in Fig. 4. These findings were generated from a simulated spherical Pt nanoparticle, utilizing the multislice method, and possessing a diameter of 11 unit cells in the [100] zone axis orientation. The distortion and noise parameters were consistent with those outlined in the preceding example. Atom counts were obtained by matching retrieved SCS values against simulated library values[51]. The improvement in column position measurements over all dose settings again indicates the proper correction of scan-line and fast-scan distortions. The improvement of SCS measurement accuracies, especially at low-dose conditions greatly decreases the chance of miscounting atoms in the structure, which in turn may be very beneficial, e.g., for the reconstruction of 3D information from atom counts[52,53].

Experimental image restorations

One of the main advantages of our image restoration method is that the training data is generated using realistic physical models of the noise found in various microscopy modalities, as well as for an appropriate range of values for the noise model parameters, as detailed in the Methods section. This methodology allows for the direct application of our network to experimental data, without requiring additional training for a particular specimen or microscope settings. Figure 5 illustrates the effectiveness of our approach on diverse types of random experimental microscopy
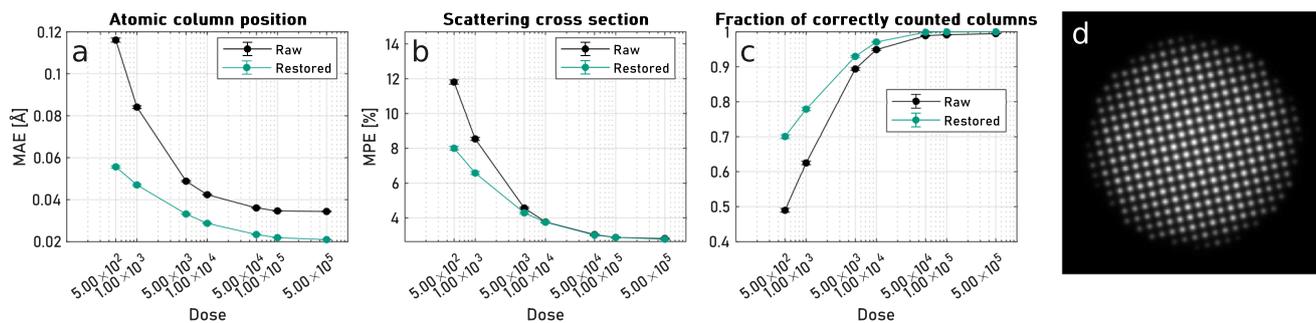
Fig. 4 Performance of the CNN restoration for atom counting. Quantification results for a spherical Pt nanoparticle with a diameter of 11 unit cells in [100] orientation. The values are based on all 333 atomic columns for 100 noise realizations. **a** The mean absolute error of the estimated atomic column positions. **b** The mean absolute percentage error of the fitted scattering cross sections, which are being used to estimate atom counts in each column. **c** The fraction of atomic columns with correctly estimated atom counts. **d** Simulated spherical Pt nanoparticle. The error bars indicate 95% confidence intervals.
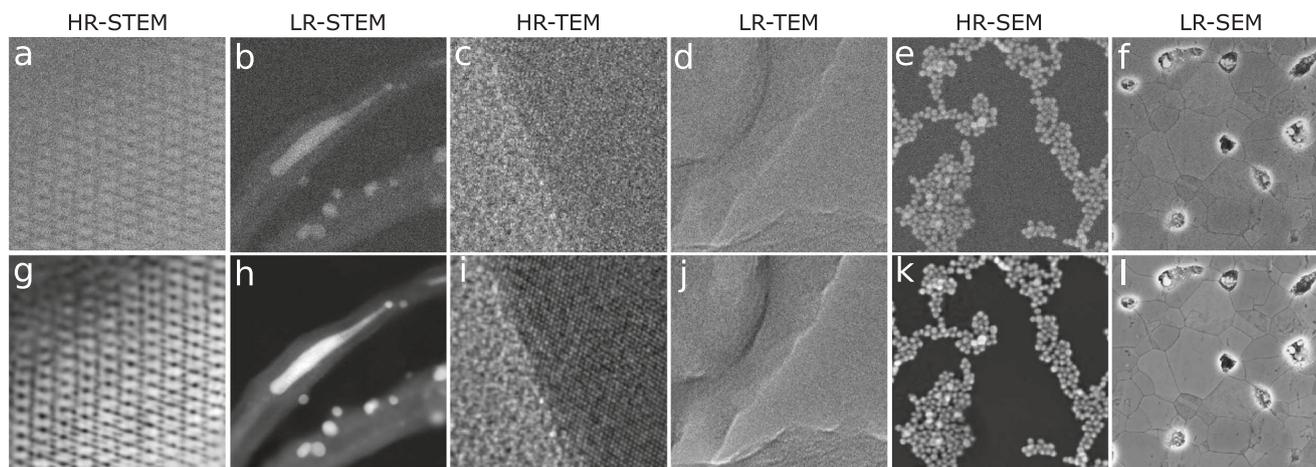

Fig. 5 Experimental image restoration for various microscopy modalities. Images (**a**)–(**f**) are the original experimental images. Images (**g**)–(**l**) are the restored versions. Images (**a**) and (**c**) are from our experimental datasets. Images (**b**) and (**d**) are taken from ref. [92], while images (**e**) and (**f**) are from ref. [93].

images. The top row of this figure shows raw experimental images for HR-STEM, LR-STEM, HR-TEM, LR-TEM, HR-SEM, and LR-SEM. The bottom row shows the corresponding restored versions of these images.

These results highlight the capability of the trained networks on experimental data and their ability to effectively handle a wide range of microscopy images with varying resolution and noise levels. Although we initially conducted only a qualitative comparison, we have subsequently carried out a more rigorous quantitative analysis between the raw and the restored data. Our first point of comparison focuses on data depicting periodic features, as shown in Fig. 5a–c. Fourier Transform analysis reveals that our network enhances the information content in these images. The second point of comparison concerns SEM data containing several particles, as illustrated in Fig. 5e. Our results show that a standard particle analysis algorithm can greatly benefit in the segmentation process when the images are pre-processed with our network. This enables a more accurate count of the particles detected in the image. Further details concerning the performance of our network on these images can be found in the Supplementary Information. It is important to note that in this study, "high resolution" refers to images with round and symmetrical features, while "low resolution" refers to images with a variety of different features. Additional examples of restored experimental images for each microscopy modality can be found in the GitHub repository https://github.com/Ivanlh20/tk_r_em.

Figure 6 illustrates the performance of our HR-STEM network when faced with a variety of experimental distribution shifts, which were not adequately considered in the initial dataset utilized for training our models. This inadequacy stems from the use of imprecise physical noise models and/or the incorrect selection of suitable parameter ranges for generating distorted data[54,55]. Figure 6a, b shows two experimental STEM images that were acquired using a Fei *Titan³ᵀᴹ* S/TEM microscope. The images were obtained using fast scanning with dwell times of 0.2 and 0.05 μs, respectively. The importance of accurately modeling fast scan distortion is evident from Fig. 6f, g. In these figures, our network architecture was trained using a model, which was not sufficient to completely compensate for the spread of pixel intensities along the scanning direction (see Eq. (48) in the Methods section). If the dwell time decreases, these image artifacts become more pronounced, as shown in Fig. 6g. While the manufacturer recommends using dwell times larger than 0.5 μs to avoid image artifacts, correctly modeling fast scan distortion allows us to fully compensate for these artifacts, as shown in Fig. 6k, l. The study of beam-sensitive materials and dynamic imaging will greatly benefit from the compensation of this distortion. Figure 6c shows a registered STEM image that contains interpolation noise. The interpolation process changes the dominant noise distribution, which can impact the restoration process, especially at low doses, as shown in Fig. 6h where some atomic columns appear blurred. However, this issue can be addressed by including this type of noise in the training dataset,
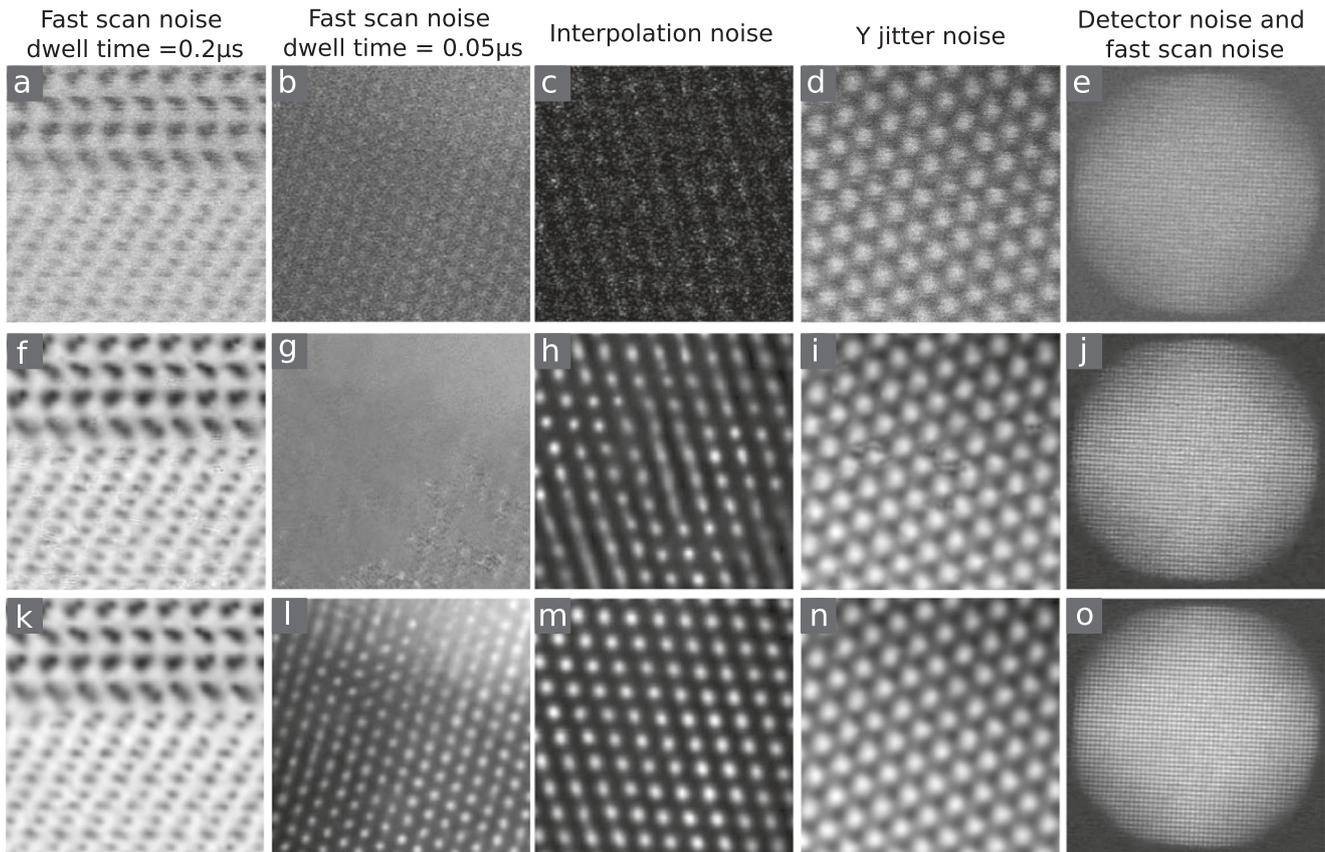
Fig. 6 **Raw STEM images alongside the results of a restoration process employing inaccurate and accurate models of the noise.** Images (**a**)–(**e**) are the original experimental STEM images. Images (**f**)–(**j**) are the restored versions of the images trained with distorted data based on inaccurate noise models. Images (**k**)–(**o**) are the restored versions of the images trained with distorted data based on accurate noise models. Images (**a**)–(**c**) were obtained from our experimental datasets, whereas (**d**) and (**e**) were obtained from refs. [94] and [56], respectively.

as explained in the Methods section. The effect of including this noise in the training dataset on the restored image can be seen in Fig. 6m, where all atomic columns become clearly visible.

Figure 6d exhibits a STEM image with strong Y-jitter distortion. The impact of an incorrect range of values for this distortion during data generation on the restored image can be seen in Fig. 6i, where some atomic columns appear split. After retraining the data with newly generated data containing the proper range of Y-jitter distortion, the neural network can correctly compensate for this image artifact, as shown in Fig. 6n. In Fig. 6e, an experimental STEM image of a nanoparticle taken using a gas cell holder is shown[56]. The dominant sources of noise in this image are detector noise and fast scan noise. Figure 6j shows a restored STEM image produced by our network architecture that was trained using a dataset generated with Shot noise as the only source of STEM detector noise (as described by Eq. (45) in the Methods section). However, this restored image exhibits strong artifacts despite using an accurate model for fast scan noise (as described by Eq. (47) in the Methods section). After retraining our network architecture with a new dataset that includes the correct STEM detector noise (as described by Eq. (46) in the Methods section), the restored image in Fig. 6o shows a significant reduction in artifacts. Nonetheless, it is worth mentioning that some of the remaining artifacts in the image could be attributed to other sources of distortion not accounted for in our data modeling, such as the gas holder effect, charging artifacts, and residual electronic noise.

Another example that highlights the importance of properly modeling noise and distortion sources can be seen in Fig. 7. In this figure, we compare the reconstruction performance of our CNN, AtomSegNet[33], and Noise2Void-NN (N2V)[57], which was retrained

on the presented experimental image itself. The sample is a $BaHfO_3$ nanoparticle (Fig. 7-③) embedded in a superconducting $REBa_2Cu_3O_{7-\delta}$ (REBCO) matrix[58,59] (Fig. 7-②), which was grown on a $SrTiO_3$ substrate (Fig. 7-①). While all three networks successfully remove the noise from the image, there are notable differences in the reconstruction results. In region①, the N2V reconstruction recovers all the weaker intensities of the $Ti + O$ columns to some degree, which is not the case for the AtomSegNet reconstruction. There, some of the columns blur or even disappear. Our CNN reliably recovers all atomic columns with superior contrast to the other two methods. Similar improvements are evident also in region② but most notably in region③. This region at the top of the image is also degraded, presumably by either FIB damage or carbon contamination. In both N2V and AtomSegNet reconstructions, features tend to blur into diagonal streaks, while our CNN recovers clearly distinguishable atomic columns and, given that the $BaHfO_3$ nanoparticle grew epitaxially on the $SrTiO_3$ substrate, that is indeed what would be expected[60]. Considering the N2V network is a generic denoising network, the results are quite remarkable, albeit the additional training step is somewhat inconvenient from a user perspective.

These results illustrate that the network presented in this work is accurate not only in terms of perceived contrast enhancement but also quantitatively, thereby boosting the accuracy and precision of atomic structure determination in ADF-STEM studies. This superior performance, compared to other networks, arises not just from the latest advances in deep learning but also from the development of accurate and physically meaningful models that account for all distortions specific to HAADF-STEM.
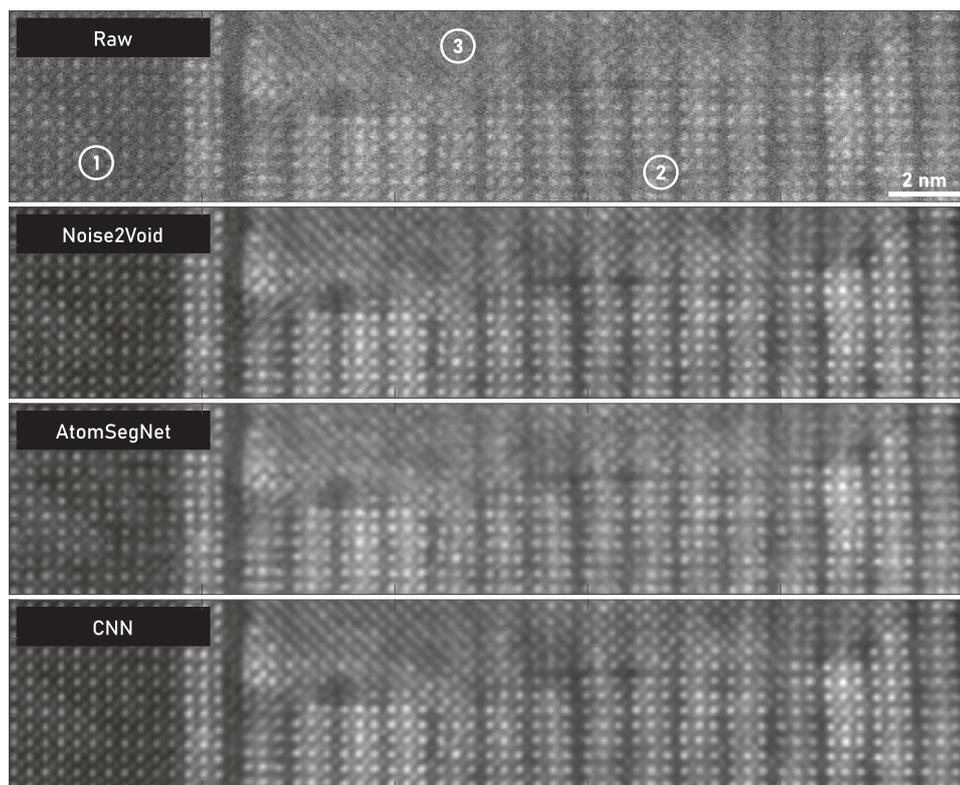
Fig. 7 Comparison of different CNN-restoration approaches on an experimental HAADF-STEM dataset. The results are shown for a $BaHfO_3$ nanoparticle (③) embedded in a superconducting $REBa_2Cu_3O_{7-\delta}$ (REBCO) matrix (②), which was epitaxially grown on a $SrTiO_3$ substrate(①). Images were acquired on a non-probe-corrected Titan microscope with 300 keV at KIT Karlsruhe. The data are described in detail in refs. [58] and [59].

Although our networks have demonstrated impressive results, it is crucial to be aware of potential pitfalls that could lead to incorrect scientific conclusions. Such pitfalls may arise when the network is applied to data that fall outside the manifold encompassed by the training data. Indeed, this situation can significantly compromise the quality of the restored data by introducing unwanted artifacts. These artifacts could, in turn, adversely affect the interpretation of the results, as elaborated in the Supplementary Information.

In addition, our research indicates that while the neural network effectively restores signals in the presence of a mixture of correlated and uncorrelated noise, it encounters challenges when uncorrelated noise greatly exceeds the recorded signal. Such high level of uncorrelated noise can lead to the false detection of atomic features, a scenario exemplified in Supplementary Fig. 2, where the recorded data is affected by Gaussian noise. The network, misinterpreting these patterns as real features, tries to eliminate the noise, which can result in the production of hallucinatory artifacts. However, this problem is less common with correlated noise because it includes signal information that the network can decode, learning from its intricate relationship with the data.

METHODS

In single-shot EM image restoration, the goal is to estimate an undistorted image y from a distorted image x. To achieve this, we train a generator G using a deep neural network approach, which learns to estimate the corresponding undistorted image y for a given input x. During the training procedure, a loss function is minimized to evaluate the quality of the results.

Traditionally, pixel-wise losses such as $\mathscr{L}_1$ or $\mathscr{L}_2$ have been used to obtain quantitative results for the image restoration problem[61]. However, these losses often lead to blurred images that do not look realistic. To address this, we propose a conditional generative adversarial network (cGAN) that trains both a generator and a discriminator. The generator G maps the distorted image x to the undistorted image $y_g = G(x)$, and the discriminator is trained to differentiate between real and generated images[62]. We use pixel-wise losses to ensure quantitative results while restricting the GAN discriminator to model high-frequency details, resulting in sharper and more realistic restored images.

Our training is supervised, requiring input pairs of distorted and undistorted EM images. However, in practice, we only have access to distorted EM data. We can partially overcome this limitation by collecting time series EM images and generating an undistorted image through an averaging procedure, based on both rigid and non-rigid registration. Nevertheless, the combination of high-speed scans, jitter, and low-dose conditions results in highly correlated distortions[13], which compromise the reliability of the averaging process. In addition, prolonged exposure to the electron beam can lead to charging, beam damage, atom hopping, and rotation of the specimen under study, thereby further undermining the averaging procedure. Consequently, the only viable solution is to train the GAN using synthetic pairs of undistorted and distorted EM images.

Network architecture

A GAN[40] is a powerful framework that encourages predictions to be realistic and thus to be close to the undistorted data distribution. A GAN consists of a generator (G) and discriminator (D) playing an adversarial game. A generator learns to produce output that looks realistic to the discriminator, while a discriminator learns to distinguish between real and generated data. The models are trained together in an adversarial manner such that
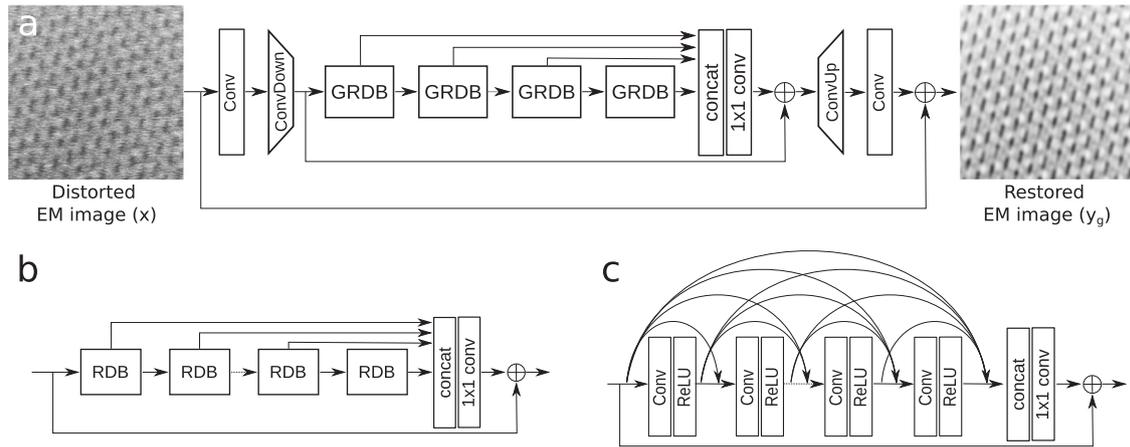
Fig. 8 Concatenated Grouped Residual Dense Network (CGRDN) architecture for EM image restoration. **a** Overall architecture, **b** GRDB architecture used in (**a**), **c** RDB architecture used in (**b**).
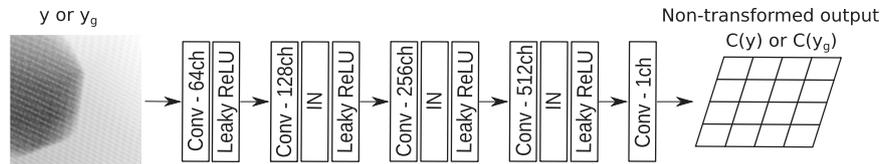

Fig. 9 Patch discriminator architecture. The result of the network is the non-transformed output $C(y)$ or $C(y_g)$.

improvements in the discriminator come at the cost of a reduced capability of the generator and vice versa. The GAN involves the generation of conditional data, which is fed to the generator and/or the discriminator[36]. The generator and discriminator architectures proposed here are adapted from those described in refs. [63] and [62], respectively. The details of these architectures are discussed in the following sections.

Generator architecture. Our generator architecture, called Concatenated Grouped Residual Dense Network (CGRDN), is shown in Fig. 8. This network architecture is an extension of the GRDN for image denoising[23], which was ranked first for real image denoising in terms of the PSNR and the structure similarity index measure in the NTIRE2019 Image Denoising Challenge[64]. The GRDB architecture is shown in Fig. 8a. The building module of this architecture is the residual dense block (RDB)[63], which is shown in Fig. 8b. The original GRDN architecture can be conceptually divided into three parts. The first part consists of a convolutional layer followed by a downsampling layer based on a convolutional stride, the middle part is built by cascading GRDBs and the last part consists of an upsampling layer based on transposed convolution followed by a convolutional block attention module (CBAM)[65] and a convolutional layer. The GRDN also includes the global residual connection between the input and the last convolutional layer. In the original version of the GRDN[23], residual connections are applied in three different levels (global residual connection, semi-global residual connection in GRDB, and local residual connection in each RDB). However, in the version submitted for the NTIRE2019 Image Denoising Challenge[64], residual connections for every 2 GRDBs were included.

Although it has been demonstrated that one architecture developed for a certain image restoration task also performs well for other restoration tasks[62,63,66,67], an architecture for a given task will be data dependent. When applied to EM data, we found out that 2 modifications of GRDN are necessary in order to best handle the nature of our data, which involves different types and levels of distortions with high correlation between pixels:

1. The cascading of the GRDN is replaced by feature concatenation, feature fusion, and a semi-global residual connection. This allows us to exploit hierarchical features in a global way, which is important for highly correlated pixels that extend over a large area of the image.
2. The CBAM, which is included in ref. [63] is removed from our network. The reason for this is the use of large image sizes (256×256) for training, which reduces its gain[23].

Discriminator architecture. The purpose of the discriminator network is to judge the quality of the output data resulting from the generator network. For our discriminator, we use the 70×70 convolutional patch discriminator described in ref. [62] with some minor modifications. The zero-padding layers were removed and batch normalization layers[29] were replaced by instance normalization layers (IN)[68]. Figure 9 shows the structure of the discriminator network. The result of the network is the non-transformed output $C(y)$ or $C(y_g)$ of dimensions 32×32.

Some benefits of the discriminator architecture shown in Fig. 9 include that it is fully convolutional and it only penalizes structure at the scale of image patches. Furthermore, we enhance our discriminator based on the relativistic GAN, which has been shown to improve the data quality and stability of GANs at no computational cost[69]. Different from the standard discriminator, which estimates the probability that input data is real, a relativistic discriminator predicts the probability that real data y is relatively more realistic than generated data $y_g = G(x)$. If we denote our relativistic average patch discriminator as $D_{Rap}(x)$, then the output of the discriminator can be written as:

$$D_{\text{Rap}}\left(y, y_g\right) = \sigma\left(C(y) - \mathbb{E}_{y_g}\{C(y_g)\}\right) \tag{2}$$

$$D_{\text{Rap}}\left(y_g, y\right) = \sigma\left(C(y_g) - \mathbb{E}_{y}\{C(y)\}\right) \tag{3}$$

where σ is the sigmoid function and $\mathbb{E}_{x_1,\ldots x_n}\{.\}$ is an operator representing the expectation value computed on the variables
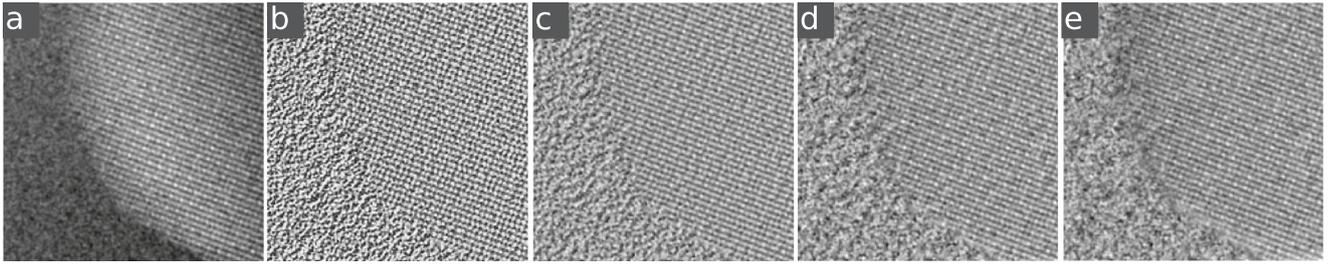
Fig. 10 Calculation of the whitening transform for different window sizes. a Undistorted ADF-STEM image of a nanoparticle on a carbon support. Images are generated by applying the whitening transform to (**a**) by using different window sizes of **b** 2, **c** 4, **d** 8 and **e** 16.

$x_1, \ldots x_n$. In the next section, these functions will be used in the definition of the loss functions.

Loss function

The loss function serves as the effective driver of the network's learning. Its objective is to map a set of parameter values of the network onto a scalar value, thereby enabling the ranking and comparison of candidate solutions. Ideally, one would directly incorporate the physics that govern the distribution of noise-free EM images into the network's architecture. However, achieving this integration has proven nontrivial, and, to date, no successful method for such an incorporation has been identified. As an alternative, we have opted to embed the physics governing this distribution into our meticulously designed loss functions. In our case, both the discriminator and adversarial losses are founded upon the relativistic average GAN loss, as defined in ref. [69]. The various contributions of these loss functions are detailed in the subsequent sections.

$\mathcal{L}_1$ *loss.* Pixel-wise losses are advantageous to keep quantitative information of the ground-truth image. In this work, we used the $\mathcal{L}_1$ loss, which as compared to the $\mathcal{L}_2$ loss yields less blurred results[61]. The $\mathcal{L}_1$ loss can be written as:

$$\mathcal{L}_1 = \mathbb{E}_{y, y_g}\{w_y \|y - y_g\|\}, \tag{4}$$

$$w_y = 1 / \max(\sigma_{\min}, \text{Std}_y\{y\}) \tag{5}$$

where w_y is a weighting factor that gives equal importance to each example regardless of its contrast, $\sigma_{\min}$ is a small value to limit the maximum scaling factor, and $\text{Std}_{x_1, \ldots x_n}\{.\}$ is an operator that represents the standard deviation calculated on the variables $x_1, \ldots x_n$.

$\mathcal{L}_2$ *loss.* Due to the design of our architecture, which is learning the residual difference between the distorted and undistorted image and based on the fact that distorted images can have few outliers in the distribution of pixel intensities (i.e. X-rays hitting the EM detector, saturation of the detector, low dose and dead-pixels), the output of the generator will show a strong correlation at those pixel positions. For this reason, we also used the $\mathcal{L}_2$ loss which strongly penalized the outliers:

$$\mathcal{L}_2 = \mathbb{E}_{y, y_g}\{w_y \|y - y_g\|^2\} \tag{6}$$

Multi-local whitening transform loss. Local contrast normalization (LCN) is a method that normalizes the image on local patches on a pixel basis[70]. A special case of this method is the whitening transform which is obtained by subtracting the mean and dividing by the standard deviation of a neighborhood from a particular pixel:

$$y_{ij}^S = \left(y_{ij} - \mathbb{E}_{\hat{S}}\{y_{i,j}\}\right) / \max\left(\sigma_{\min}, \text{Std}_{\hat{S}}\{y_{i,j}\}\right), \tag{7}$$

where $\hat{S}$ is a local neighborhood around the pixel i, j of window size S. The whitening transform makes the image patches less correlated with each other and can highlight image features that were hidden in the raw image due to its low local contrast. This effect can be seen in Fig. 10a, which shows a simulated ADF-STEM image of a random nanoparticle on a carbon support. The edge of the nanoparticle shows low contrast due to its reduced thickness, resulting in lower intensity values. Based on this observation, we introduce a multi-local whitening transform (MLWT) loss, which pays more attention to fine details independent of the intensity value. Specifically, the generated and the ground-truth image are local whitening transforms corresponding to different window sizes of 2×2, 4×4, 8×8, and 16×16 pixels.

Using different window sizes for the calculation of the whitening transform, we ensure that the relevant features present in the image are highlighted independently of its pixel size. Figure 10b–e shows an enhancement of the edge of the nanoparticle as well as the carbon support after applying the whitening transform to Fig. 10a by using different window sizes.

Then, we calculate the average $\mathcal{L}_1$ loss for these four images:

$$\mathcal{L}_{\text{mlwt}} = \frac{1}{4} \sum_{S=2,4,8,16} \mathbb{E}_{y^S, y_g^S}\{\|y^S - y_g^S\|\}. \tag{8}$$

Fourier space loss. In electron microscopy, Fourier space contains crucial information about the sample and any distortions that may be difficult to discern in real space. To address this issue, we introduce the $\mathcal{L}_\gamma$ loss in the 2D Fourier transform of the difference between the generated data y_g and the ground-truth image y. Nevertheless, it is noted that high-frequency information typically possesses smaller values than low-frequency information. Consequently, to accentuate the high-frequency information, we apply a power transform to the aforementioned difference and define the loss function as follows:

$$\mathcal{L}_{\text{fs-}\gamma} = \mathbb{E}y, y_g\left[|\mathcal{F}(y - y_g)|^\gamma\right], \tag{9}$$

Here, $\mathcal{F}$ symbolizes the 2D Fourier transform, and γ is a parameter in the range $(0.0, 1.0)$. In our investigation, we utilize $\gamma = 0.125$.

Constraint losses. Some important parameters for EM quantification are the total intensity and the standard deviation of the images. The reason for this is that they carry information about physical quantities of the sample or microscope, such as the number of atoms, defocus and spatial and temporal incoherence[71,72]. Therefore, we encourage that the restored images have to minimize the above quantities, resulting in the following two loss functions:

$$\mathcal{L}_{\text{mean}} = \|\mathbb{E}_y\{y\} - \mathbb{E}_{y_g}\{y_g\}\|, \tag{10}$$

$$\mathcal{L}_{\text{std}} = \|\text{Std}_y\{y\} - \text{Std}_{y_g}\{y_g\}\|. \tag{11}$$

Adversarial loss. The job of the relativistic adversarial loss is to fool the discriminator, which can be expressed as:

$$\mathcal{L}_{\text{Adv}} = -\mathbb{E}_{x,y}\left\{\log\left(1 - D_{\text{Rap}}(y, y_g)\right)\right\} - \mathbb{E}_{y_g}\left\{\log\left(D_{\text{Rap}}(y_g, y)\right)\right\}, \quad (12)$$

with $D_{\text{Rap}}(y, y_g)$ and $D_{\text{Rap}}(y_g, y)$ defined in Eqs. (2) and (3), respectively. This definition is based on the binary cross entropy between the ground-truth and the generated images. Different from the conventional adversarial loss, in which y is not used, our generator benefits from y and y_g in the adversarial training.

Generator loss. Our total generator loss function can be written as:

$$\mathcal{L}_G = \mathcal{L}_{\text{pixel-wise}} + \lambda_{\text{Adv}}\mathcal{L}_{\text{Adv}}, \quad (13)$$

$$\mathcal{L}_{\text{pixel-wise}} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 + \lambda_{\text{mlwt}}\mathcal{L}_{\text{mlwt}} + \lambda_{fs-\gamma}\mathcal{L}_{fs-\gamma} \\ + \lambda_{\text{mean}}\mathcal{L}_{\text{mean}} + \lambda_{\text{std}}\mathcal{L}_{\text{std}}, \quad (14)$$

where $\mathcal{L}_{pixel-wise}$ is our pixel-wise loss function, $\lambda_1, \lambda_2, \lambda_{\text{mlwt}}, \lambda_{fs-\gamma}, \lambda_{\text{mean}}, \lambda_{\text{std}}$ and λ_{Adv} are the weighting parameters to balance the different loss terms.

Discriminator loss. Symmetrically to the relativistic adversarial loss, the relativistic discriminator is trying to predict the probability that real data is relatively more realistic than generated data, and it can be expressed as:

$$\mathcal{L}_D = -\mathbb{E}_{x,y}\left\{\log\left(D_{\text{Rap}}(x, y, y_g)\right)\right\} - \mathbb{E}_{x,y_g}\left\{\log\left(1 - D_{\text{Rap}}(x, y_g, y)\right)\right\}. \quad (15)$$

Data generation

While it is possible to fully describe the electron-specimen interaction and image formation in an electron microscope, generating realistic EM image simulations for specimens on support with sizes of a few nanometers is too time-consuming even with the most powerful GPU implementations of the multislice method[47,48]. However, our goal is to train a neural network to correct EM distortions without the need to know the specific specimen or microscope settings. Therefore, we only need to generate undistorted images that closely mimic the appearance of real EM data, while the EM distortions must be accurately modeled. The generated undistorted images should also incorporate the physical parameters of both the specimen and the microscope settings. For high-resolution features, the foundation is an atomic object, the parameters of which may encompass atomic sizes, atomic distances, atomic vibrations, lattice parameters, and relative intensities of atomic species. Conversely, when generating low-resolution images, the building blocks are points, to which general objects can be assigned. In addition, the generated undistorted images should be linked to parameters such as acceleration voltage, aberrations, magnification, detector sensitivity, detector angles, and the transfer function of the detection system.

Specimen generation. In order to optimize the simulation process, we generate a specimen that fully covers the extended simulated box size $\hat{l}^e_{xyz}$. This is an expanded version of the required simulation box size $\hat{l}_{xyz}$. The calculation of $\hat{l}_{xyz}$ starts by randomly selecting a pixel size dr within the range $[0.025, 0.90]$Å. This range is chosen specifically to capture most of the high-resolution features typically observed in EM images. When low-resolution features are being created, the same simulation box size is retained primarily for optimization purposes. However, relevant parameters are adjusted to ensure a consistent pixel size range. By using the required image size (n_x, n_y), $n_z = \max(n_x, n_y)$ and dr, the required simulation box size can be expressed as $\hat{l}_{xyz} = \{n_x dr, n_y dr, n_z dr\}$. From these values, an extended number of pixels $n^e_i = n_i + \text{round}(d_{ext}/dr)$ and an extended simulation

box size $\hat{l}^e_{xyz} = \{n^e_x dr, n^e_y dr, n^e_z dr\}$ are obtained, where d_{ext} is the maximum correlation distance for a given value of scanning distortions. The specimen generation is divided in 3 steps.

The first step of specimen generation involves randomly selecting a specimen type from the following options: crystalline specimen, amorphous specimen, or individual points. If the selected specimen is crystalline, the generation process starts by randomly choosing up to 16 unique atomic types with atomic number Z in the range $[1, 103]$. The crystallographic space group is randomly chosen from a range $[1, 230]$. The lattice parameters and the angles of the chosen space group are selected randomly from a range $[3.1, 25.0]$Å and $[45°, 120°]$, respectively. Atomic positions of the asymmetric unit cells are generated randomly within the volume that is allowed by their space-group symmetry. This specimen generation process is subject to a physical constraint: after applying the space group symmetry to the atomic positions on the asymmetric unit cells, the minimum distance between the atoms in the unit cell must be within the range $[0.95, 7.0]$ Å. If this requirement is not met, the generation process is restarted. The generation of amorphous specimens is based on randomly choosing only one atomic number Z from the range $[1, 103]$. The atomic positions of amorphous specimens are generated by randomly placing atoms within the extended simulation box, subject to the requirement that the minimum distance between atoms is within the range $[0.95, 1.6]$ Å. This process continues until the desired density within the range $[2.0, 7.0]$ g/cm^3 is achieved. In contrast, the generation of individual points starts by randomly choosing a number of points within a given range of positive integers. The 3D positions of the particles are then generated randomly within the extended simulation box, subject to the requirement that the minimum distance between particles is within the range $[1, 20]$dr. This option is also used to generate low-resolution images.

The second step begins by randomly choosing between a specimen orientation along the zone axis or a random orientation. The probability of choosing a zone axis orientation is 0.75. If the specimen is crystalline, the zone axis orientation is randomly chosen from the first eight main zone axes, and a small random mistilt angle is generated for the chosen orientation using a normally distributed random number with a standard deviation of 5°. For non-crystalline specimens, a random 3D orientation is generated. To prevent alignment of crystalline specimens along the xy directions, an additional random rotation is applied along the z-axis. For a given generated orientation, the specimen is oriented and cropped in the xy plane so that it fits within the extended simulated box. This is followed by a random generation of a wedge on the specimen with a probability of 0.75. The wedge can be generated on the top, bottom, or both surfaces of the specimen, each with a probability of occurrence of 0.33. The wedge orientation is generated randomly in the xy plane, and its angle is chosen randomly from the range $[5°, 45°]$. Shapes can be applied to the specimen with a probability of 0.5. To avoid any preference for the three different types of shapes, the probability of occurrence for each type is set to 0.33. The first type of shape is a polygon rod, for which the number of cross-section vertices sliced along its length is randomly chosen from the range $[3, 15]$. The rod is also placed and oriented randomly. The radius of the polygon is chosen randomly from the range $[0.01, 0.5]\max(\hat{l}_{xyz})$. The second shape is a convex polyhedron, for which the radius and the number of vertices are chosen randomly from the ranges $[0.01, 0.5]\max(\hat{l}_{xyz})$ and $[4, 20]$, respectively. The third shape is a hard shape, in which all atoms on one side of a randomly generated 3d plane parallel to the z orientation are removed. The application of a chosen shape can be used to either remove or keep the atoms of the specimen, with a probability of keeping the atoms of 0.5. Defects are generated randomly with a probability of 0.8. The process starts by randomly selecting a number of atoms, n_{sel}, within the specimen. This number is chosen randomly from

the range $[0, n_{max}]$, where n_{max} is equal to the number of atoms in the specimen multiplied by 0.25 and rounded to the nearest whole number. The positions of the selected atoms are randomly changed with a probability of 0.5. This is done by adding a normally distributed random number with a standard deviation equal to the atomic radius to the position of each selected atom.

The final step of specimen generation adds a support layer with a probability of 0.95. The support layer can be either crystalline or amorphous, each with a probability of 0.5. The thickness of the support layer is chosen randomly from the range $[1, 30]$nm. The process described above for crystalline and amorphous specimen generation is used for the support layer, with the exception of shape generation. Finally, the generated atoms are added to the specimen.

Undistorted data generation. High/medium resolution electron microscopy data can be synthesized as a linear superposition of the projected signal of each atom in the specimen at a given orientation. Moreover, each projected atomic signal can be modeled as a two-dimensional radial symmetric function, $f_Z^i(r)$, where the index i refers to an atom with atomic number Z in the specimen. Under this assumption, y can be expressed as:

$$y = \sum_Z \sum_i f_Z^i(|\mathbf{r} - \mathbf{r}_i|), \tag{16}$$

where $\mathbf{r}$ is a two-dimensional vector. In addition, we model $f_Z(r)$ for each atom with atomic number Z as a weighted sum of Gaussian, Exponential, and Butterworth functions:

$$f_Z(r) = h_1 e^{-\frac{r^2}{2\left(r_Z^m\right)^2}} + h_2 e^{-\frac{r}{r_Z^m}} + \frac{h_3}{1 + (r/r_Z^m)^{2n}}, \tag{17}$$

where h_1, h_2, h_3, n and r_m are the parameters of our model, which are restricted to positive values. This parameterization has three benefits. First, it accurately models almost any simulated/experimental incoherent EM image. Second, it allows for an easy inclusion of physical constraints. Third, it only requires five parameters. To allow realistic tails of $f_Z(r)$, we constrain n to be a uniform random variable between $[4.0, 16.0]$. We would also like to emphasize that all numerical ranges for the data generation were fine-tuned based on analyzing around 2000 real simulations of (S)TEM images for different specimens and microscope settings.

In order to encode physical information into this model, r_Z^m is chosen proportionally to the transformed two-dimensional mean square radius, $\hat{r}_Z$, of the projected atomic potential, $V_Z^p(r)$[73]:

$$r_Z^m = a \times (\hat{r}_Z)^\alpha + b \tag{18}$$

where

$$a = \text{Std}_Z\{\hat{r}_Z\}/\text{Std}_Z\{(\hat{r}_Z)^\alpha\}, \tag{19}$$

$$b = \mathbb{E}_Z\{\hat{r}_Z\} - a \times \mathbb{E}_Z\{(\hat{r}_Z)^\alpha\}, \tag{20}$$

$$\hat{r}_Z = \left[\frac{\int_0^\infty r^2 V_Z^p(r) r \, dr}{\int_0^\infty V_Z^p(r) r \, dr}\right]^{1/2} \tag{21}$$

and α is a uniform random variable between $[0.75, 1.25]$. On the other hand, the linear coefficients h_1, h_2 and h_3 are randomly chosen within the range $[0.5, 1.0]$ with the following constraint:

$$\int f_{Z_i}(r) dr > \int f_{Z_j}(r) dr, \text{ if } Z_i > Z_j \tag{22}$$

where Z_i and Z_j are the atomic numbers of two elements of the specimen. This constraint arises from the fact that the integrated intensity of quasi-incoherently scattered electrons of a given atomic number is proportional to Z^γ, in which γ is a real number between 1.0 and 2.0 depending on the microscope settings[74].

The process of generating low-resolution images begins by randomly choosing a set of low-resolution image types from the following options: soft particles, sharp particles, grains, bands, boxes, and cracks. This stage uses the specimen type "individual points" to generate random positions where different objects will be placed. Finally, the low-resolution image is obtained by linearly superimposing these individual objects.

The generation of soft particles starts by randomly choosing a number of particles in the range $[15, 85]$. Each soft particle image is generated by randomly rotating the asymmetric version of Eq. (17), where $r_Z^m = (r_Z^{m_x}, r_Z^{m_y})$ and $r_Z^{m_y} = a r_Z^{m_x}$, with a a random variable in the range $[0.8, 1.2]$. In the case of sharp particles, there is a sharp transition between the border and background of the particle, and the particle can be either polygonal or elliptical with equal probabilities of occurrence. The process starts by randomly choosing a number of particles in the range $[15, 40]$. For the polygon option, the number of vertices is randomly chosen in the range $[3, 5]$. Each sharp particle image is generated by masking a 3D random positive plane intensity with its randomly rotated shape. This masking creates an intensity gradient over the $x - y$ plane such that the object does not appear flat.

Grain generation in $2D$ is performed using the Voronoi tessellation method[75], which is one of the available techniques for producing random polygonal grains within a domain. This process starts by randomly selecting a number of points within the range $[15, 175]$. Each grain image is created by masking a 3D random positive plane with its corresponding Voronoi cell. In addition, the grain borderline is included with a probability of occurrence of 0.5, where its intensity value is randomly assigned within the range $[0.5, 1.5] \times \text{mean}$ (grainintensity).

EM images may exhibit contrast inversion related to the projected specimen, which can be easily simulated by inverting the image:

$$y \leftarrow \max(y) - y. \tag{23}$$

The probability of this mechanism occurring was set to 0.5. To introduce nonlinear dependence between the generated image intensity and the projected specimen's structure, y is non-linearly transformed with a probability of occurrence of 0.5:

$$y \leftarrow |y|^\beta \tag{24}$$

where β is a uniform random number selected from the range $[0.5, 1.5]$.

To further break this linearity, a random background was added to y. The background is randomly chosen between a 3D plane and a Gaussian, with an occurrence probability of 0.5 for each. In the first case, a randomly orientated positive 3D plane is generated with a random height between $[0, \max(y)/2]$. In the second case, the Gaussian center and its standard deviation are randomly chosen within the range of the xy simulation box size and $[0.2, 0.6] \times \min(n_x, n_y)$, respectively. From the analysis of the experimental and simulated data, we found that the ratio $r_{std/mean} = \text{Std}\{y\}/\mathbb{E}\{y\}$ is between $[0.01, 0.35]$. Therefore, if the EM image does not fulfill the latter constraint, then it is linearly transformed as:

$$y \leftarrow cy + d \tag{25}$$

where c and d are chosen to bring $r_{std/mean}$ within the range of the constraint. Finally, the EM image is normalized through dividing by its maximum value.

$$y \leftarrow \frac{y}{\max(y)} \tag{26}$$

Note that the correct parameterization of the model and the randomness of its parameters are subject to physical constraints allowing to encode information in the generated high/medium resolution EM image of the atomic size, atomic vibration, relative intensities between atomic species, detector angle, acceleration voltage, aberrations and/or detector sensitivity.

TEM noise model

The TEM noise model is based on the fact that TEM images are recorded using parallel illumination, and that most signal acquisitions for electrons are set up so that the detector output is directly proportional to the time-averaged flux of electrons reaching the detector. In case of TEM, the electrons are detected indirectly using a charge-coupled device (CCD) sensor[76] or a complementary metal oxide semiconductor (CMOS) sensor[77], or directly using a direct electron detector[78].

For indirect detection, primary electrons are converted to photons in a scintillator, which are then directed to the CCD/CMOS sensor through a lens or fiber-optic coupling. In contrast, for direct electron detectors, the CMOS sensor is directly exposed to the electron beam.

TEM camera modulation-transfer function. Scattering of incident electrons over the detector leads to the detection of electrons in multiple pixels, which can be quantitatively described using the modulation-transfer function (MTF). Because the effect of the MTF is to produce an isotropic smear out of features on the recorded TEM image, which in general cannot be distinguished from an undistorted TEM image recorded with other microscope settings, we embedded this effect into the undistorted TEM image by convolving it with the point-spread function (PSF), which is the Fourier transform of the MTF:

$$y \leftarrow y \otimes \mathrm{PSF}. \tag{27}$$

The MTF itself can be separated into a rotationally symmetric part, MTF_r, describing the spread of electrons in the detector, and a part describing the convolution over the quadratic area of a single pixel. This yields the following equation:

$$\mathrm{MTF} = \mathrm{MTF}_r \mathrm{sinc}\,(\pi u/2)\mathrm{sinc}\,(\pi v/2), \tag{28}$$

where the Fourier space coordinates (u, v) are defined in units of the Nyquist frequency[79]. Furthermore, we found that the general shape of MTF_r can be expressed parametrically as:

$$\mathrm{MTF}_r = ae^{-\frac{g^2}{2b^2}} + (1-a)e^{-\frac{g^2}{2c^2}}, \tag{29}$$

where a, b and c are positive real numbers. These numbers were randomly generated until they fulfilled the constraint that on a numerical grid of 1000 points with a length of 10 units of the Nyquist frequency, the MTF_r is a positive and monotonically decreasing function.

TEM detector noise. TEM detectors are subject to three main sources of noise: shot noise, dark-current noise, and readout noise. These noise sources can be classified into two types: temporal and spatial noise. Temporal noise can be reduced by frame averaging, whereas spatial noise cannot. However, some spatial noise can be mitigated by using techniques such as frame subtraction or gain/offset correction. Examples of temporal noise discussed in this document include shot noise, reset noise, output amplifier noise, and dark-current shot noise. Spatial noise sources include photo-response nonuniformity and dark-current nonuniformity. Each of these noise sources can lower the SNR of a sensor imaging device.

Photon shot noise. After the initial conversion of the incident electron to its photon counterpart, the generated photons will hit the photosensor pixel area, liberating photoelectrons proportional to the light intensity. Due to the quantum nature of light, there is an intrinsic uncertainty arising from random fluctuations when photons are collected by the photosensor. This uncertainty is described by the Shot process $\mathbb{P}$ with mean ax, where a is a dose scale factor.

The distribution of a is exponential, with a scale parameter of 0.5 and a range $[0.5, 750]/\mathbb{E}\{y\}$. The use of the exponential distribution yields higher probabilities for the generation of images at lower doses which is the focus of our research. The division by a in the

equation below brings x back to its original range:

$$x \leftarrow \frac{P(ax)}{a} \tag{30}$$

Fixed-pattern noise. Fixed-pattern noise (FPN) is a pixel gain mismatch caused by spatial variations in the thickness of the scintillator, fiber-optic coupling, substrate material, CCD bias pattern, and other artifacts that produce variations in the pixel-to-pixel sensitivity and/or distortions in the optical path to the CCD or in the CCD chip itself[80]. Since FPN is a property of the sensor, it cannot be fully eliminated. However, it can be suppressed using a flat-field correction procedure. We model the remaining distortion as a normal distribution $\mathbb{N}$ with zero mean and standard deviation σ_{fpn}.

$$x \leftarrow x + x\mathbb{N}(0, \sigma_{\mathrm{fpn}}) \tag{31}$$

Dark-current noise. Dark current is the result of imperfections or impurities in the depleted bulk Si or at the SiO_2/Si interface. These sites introduce electronic states in the forbidden gap which allows the valence electrons to jump into the conduction band and be collected in the sensor wells. This noise is independent of electron/photon-induced signal, but highly dependent on device temperature due to its thermal activation process[81].

Dark-current nonuniformity. Dark-current nonuniformity (DCNU) arises from the fact that pixels in a hardware photosensor cannot be manufactured exactly the same and there will always be variations in the photodetector area that are spatially uncorrelated, surface defects at the SiO_2/Si interface, and discrete randomly distributed charge generation centers[82]. This means that different pixels produce different amounts of dark current. This manifests itself as a fixed-pattern exposure-dependent noise and can be modeled by superimposing two distributions. The Log-Normal distribution ($\ln\mathbb{N}$) is used for the main body and the uniform ($\mathbb{U}$) distribution is used for the "hot pixels" or "outliers"[83].

$$\mathrm{DCNU} \leftarrow \ln\mathbb{N}(\mu, \sigma) + \mathbb{U}(a, b) \tag{32}$$

with μ the mean value, σ the standard deviation, $a = \mu + 5\sigma$, and $b = \mu + 8\sigma$.

Dark-current shot noise. Additional noise arises from the random arrival of electrons generated as part of the dark signal, which is governed by the Shot process. To simulate a single frame, it is necessary to apply shot noise to the DCNU array.

$$x \leftarrow x + \mathbb{P}(\mathrm{DCNU}) \tag{33}$$

Readout noise. Readout noise is temporal noise and is generally defined as the combination of the remainder circuitry noise sources between the photoreceptor and the ADC circuitry. This includes thermal noise, flicker noise and reset noise[84].

Thermal noise. Thermal noise arises from equilibrium fluctuations of an electric current inside an electrical conductor due to the random thermal motion of the charge carriers. It is independent of illumination and occurs regardless of any applied voltage. The noise is commonly referred to as Johnson noise, Johnson-Nyquist noise, or simply white noise. It can be modeled by the normal distribution with zero mean and an appropriate standard deviation σ[84].

$$x \leftarrow x + \mathbb{N}(0, \sigma) \tag{34}$$

Flicker noise. Flicker noise, also known as $1/f$ noise or pink noise, is often caused by imperfect contacts between different materials at a junction, including metal-to-metal, metal-to-semiconductor,

and semiconductor-to-semiconductor. MOSFETs are used in the construction of CMOS image sensors, which tend to exhibit higher levels of $1/f$ noise than CCD sensors[82]. The amount of flicker noise in a CCD sensor depends on the pixel sampling rate. The equation below describes the effect of flicker noise on a signal x:

$$x \leftarrow x + \mathscr{F}(\mathbb{N}(0,\sigma)/f) \tag{35}$$

Here, $\mathscr{F}$ is the two-dimensional Fourier transform, σ is the appropriate standard deviation, and f is the reciprocal distance.

Reset noise. Before a measurement of the charge packet of each pixel is taken, the sense node capacitor of a specific row is reset to a reference voltage level. This causes all pixels in that row to be exposed to noise coming in through the reset line, transfer gate, or read transistor. As a result, images may have horizontal lines due to the fixed and temporal components of the noise. This type of noise, known as reset noise (RN), follows a normal distribution with mean zero and a standard deviation σ. It can be simulated by adding a random intensity value, generated for each row, to the intensity values of all pixels in that row[83]:

$$x \leftarrow x + \mathcal{N}(0,\sigma) \tag{36}$$

Black pixel noise. Black pixels are dots or small clusters of pixels on the sensor that have significantly lower response than their neighbors, resulting in black spots on the image. Some black pixels may be created during the production process of the CCD camera, while others may appear during its lifetime. Black pixels are time-invariant and will always appear at the same locations on the image. They can be modeled by generating a sensitivity mask (S_{Black}) with a spatially uniform distribution of a specified number of black points. Regions can be generated by applying a random walk process for a given number of random steps to the black point positions. The equation below describes the effect of black pixels on a signal x:

$$x \leftarrow x S_{Black} \tag{37}$$

Zinger noise. Zingers are spurious white dots or regions that can appear randomly in CCD images[85]. Electron-generated X-rays, cosmic rays, and muons can produce a burst of photons in the scintillator, resulting in white spots or streaks in the image. Radioactive elements (such as thorium) present in fiber-optic tapers

can also cause zingers[80]. They can be modeled by generating a sensitivity mask (S_{Zinger}) with a spatially uniform distribution of a specified number of zinger points. Similar to the black pixel noise, regions can be generated by applying a random walk process for a given number of steps to the zinger point positions:

$$x \leftarrow x S_{Zinger} \tag{38}$$

Upper-clip noise. Upper-clip noise, also known as saturation noise, is a type of noise that occurs when the intensity value of a pixel exceeds the maximum value that the CCD sensor can detect. This causes the pixel to be clipped at the maximum value, resulting in an overly bright image with lost details. This type of noise can be modeled by setting a threshold value for the maximum intensity and clipping any pixel values above that threshold T_u: Upper-clip noise. Upper-clip noise, also known as saturation noise, is a type of noise that occurs when the intensity value of a pixel exceeds the maximum value that the CCD sensor can detect. This causes the pixel to be clipped at the maximum value, resulting in an overly bright image with lost details. This type of noise can be modeled by setting a threshold value for the maximum intensity and clipping any pixel values above that threshold T_u:

$$x \leftarrow \min(x, T_u) \tag{39}$$

Quantization noise. To generate a digital image, the analog voltage signal readout during the last stage is quantized into discrete values using analog-to-digital conversion (ADC). This process introduces quantization noise, which can be modeled with respect to the ADC gain a:

$$x \leftarrow \text{round}(ax) \tag{40}$$

Figure 11 shows simulated TEM images with different types of noise. These distortions have been randomly added to the images to mimic real TEM conditions and make it easier to identify the different types of noise.

S(T)EM noise model

S(T)EM images are formed one pixel at a time by scanning a convergent electron beam along scan lines across the sample with
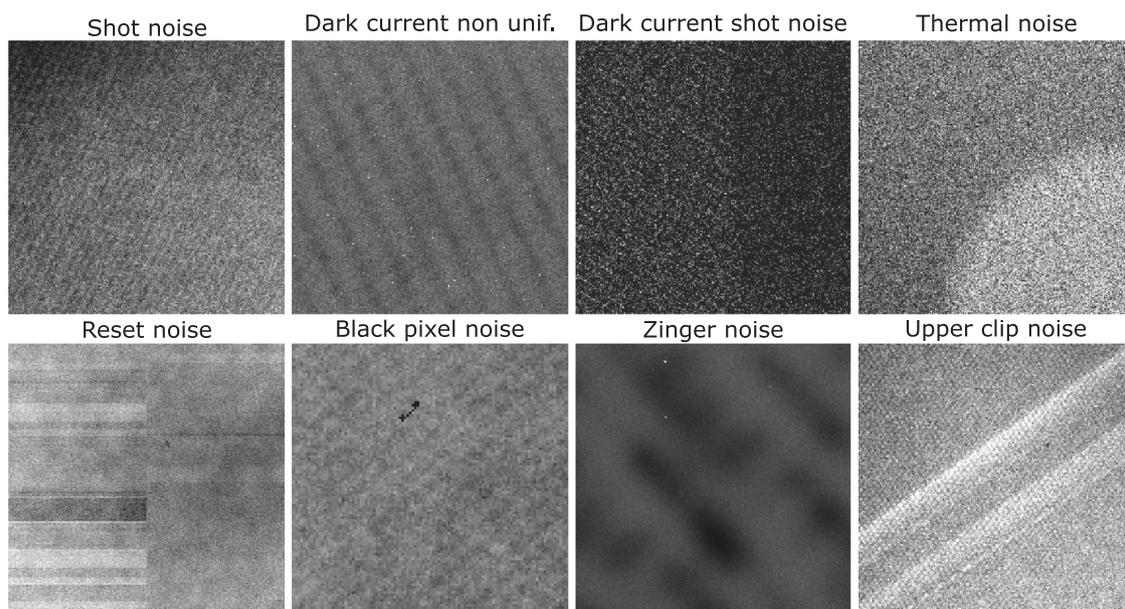

Fig. 11 Demonstration of the various types of noise in TEM. Random distorted simulated TEM images are shown for various types of noise.
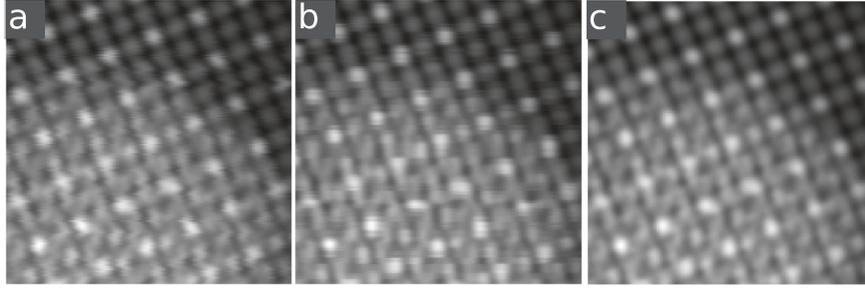
Fig. 12 Demonstration of scanning jitter distortion. Image (**a**) and (**b**) are distorted jitter images along the fast and slow scan direction, respectively. **c** Undistorted ADF-STEM image of a random sample.

constant stationary probing, which is known as dwell time. The dimension of each square-shaped pixel in the physical space is determined by the magnification. The scanning direction is called the fast/row scan direction. For conventional scan patterns, the scanning begins at the top left corner and after scanning one row of n pixels, the electron probe moves to the next row's first pixel. The time required to move the beam to the beginning of the scan line is commonly known as fly-back-time. Inaccuracies in beam positions during the scanning process give rise to characteristic scan-line/jitter distortions. Despite all technical improvements in the design of high-performance S(T)EM[3], the presence of these distortions on the recorded images still hampers the extraction of quantitative information from the sample under study[5].

Scanning jitter distortion. Scanning jitter (SJ) is caused by beam instabilities while scanning a raster pattern across the sample during the image acquisition process. There are two distinguishable jitter effects: X-jitter causes random pixel shifts along the fast-scan direction, while Y-jitter causes stretching or squishing of scan lines or line interchanges along the slow-scan direction[11]. Although these displacements are not completely random due to serial acquisition, they depend on the previous scan position. Realistic modeling of scanning jitter distortion can be achieved using the Yule–Walker correlation scheme on time series[86,87]. Furthermore, the fast and slow scanning directions can be modeled independently due to their different time scales. Here, we focus on displacement series in discrete pixels, in which each term of the series depends on the previous one. Mathematically, these displacement series can be described as:

$$\Delta_t^k = \frac{a_t^k}{\sqrt{(1-\phi_t^2)}} \text{ if } k=1$$
$$\Delta_t^k = \phi\Delta_t^{k-1} + a_t^k \text{ if } k>1 \tag{41}$$

where $t = x, y$ and k is the pixel index along a given t direction. ϕ_t is the correlation coefficient which describes the coupling between two consecutive values of the series within the range [0, 1]. a_t^i is a normally distributed random number with zero mean and standard deviation σ_t. The distorted image is created by using bicubic interpolation and evaluating on the non-regular grid, which is built by adding the positions of the regular grid and the generated displacements.

$$x \leftarrow SJ(y) \tag{42}$$

The described effects of individual jitter distortions for $\sigma_x = \sigma_y = 0.75$ and $\phi_x = \phi_y = 0.6$ along the fast and slow scan directions can be seen in Fig. 12a, b, respectively. Figure 12c shows the undistorted ADF-STEM random-generated image.

Based on our analysis of experimental data, we set the occurrence probability of jitter distortion to 0.9. In addition, we assign the occurrence probability of the X-jitter, Y-jitter and the XY-jitter to 0.25, 0.25 and 0.50, respectively. The values of σ_t and ϕ_t are randomly chosen within the range [0.0025, 0.8] Å and [0.0, 0.7], respectively.

S(T)EM detector noise. Electrons are detected by a scintillator coupled to a photomultiplier tube (PMT) via a mirror or reflective tube. Impact of the incident electrons on the scintillator cause photons to be emitted, which are directed to the PMT through a light pipe. The PMT consists of a photocathode that emits photoelectrons when illuminated by these photons, followed by a series of stages amplifying the signal. The resulting current at the anode can be measured using conventional ADC electronics[8]. The statistics of the electron multiplication as a series of Shot events with full width at half maximum (FWHM) of the pulse at the anode per single incident electron is given by[88]:

$$FWHM = 2\sqrt{2\log 2}m_c\eta G\sqrt{\frac{1-\eta+\frac{1}{\delta-1}}{m_c\eta} + \frac{\delta_c^2}{m_c^2}} \tag{43}$$

This equation assumes that the secondary gain δ at each stage inside the PMT is the same. In this equation, G represents the PMT gain, η is the detective quantum efficiency, m_c is the number of photons collected per incident electron, and δ_c^2 is the variance of that number[88]. A good approximation for the noise spectrum of a photomultiplier is the Shot distribution, which can be approximated by a Gaussian distribution for large means. Since for each electron reaching the scintillator, around 100 photons reach the cathode of the photomultiplier, a Gaussian approximation can be used with a standard deviation

$$\sigma = m_c\eta G\sqrt{\frac{1-\eta+\frac{1}{\delta-1}}{m_c\eta} + \frac{\delta_c^2}{m_c^2}} \tag{44}$$

In addition, the number of electrons hitting the scintillator is described by the Shot process ($\mathbb{P}$)[89]. The signal can therefore be constructed in two steps:

$$x \leftarrow \mathbb{P}(ax) \tag{45}$$

$$x \leftarrow (x + \mathbb{N}(0,\sigma))/a \tag{46}$$

where a is a dose scale factor. Dividing by a in the latter equation brings x back to approximately its original range.

Fast scan noise. Fast scan noise arises due to the use of short dwell times during data acquisition and appears as horizontal blur in the recorded images. This effect can also be seen in the Fourier domain as a damping effect on the high frequencies in the horizontal direction. This blurring is caused by the finite decay time of the detection system, which consists of a scintillator, a photomultiplier, and additional readout electronics[89,90]. In addition to blurring in the horizontal direction, fast scans may introduce other artifacts due to the limited response time of the scan coils. In particular, strong distortions may appear on the left-hand side of the images due to the discontinuity in the scan pattern between consecutive lines. This can be avoided by using a small delay (fly-back time) between scanning lines. The optimal value of this delay is hardware specific but results in an additional dose to the sample, which will be localized on the left-hand side of
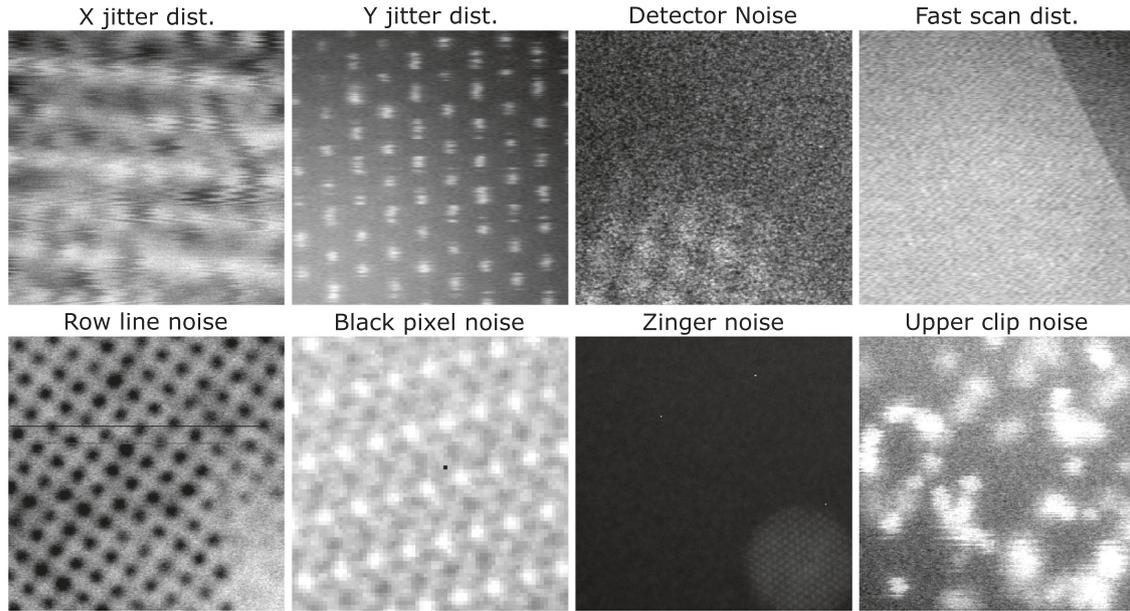
Fig. 13 Demonstration of the various types of noise in STEM. Random distorted simulated STEM images are shown for various types of noise.

each image[91]. In general, the effect of fast scan distortion can be modeled by convolution in one dimension along the fast-scan direction between x and the point-spread function (PSF) of the system. After careful analysis of the experimental data, we find that the PSF of the system can be decomposed into contributions from the detector and the readout system.

$$\mathrm{Im}_{fsd}(x, y) = \mathrm{Im} \circledast \mathrm{psf}_{detector} \circledast \mathrm{psf}_{readout} \qquad (47)$$

with

$$\mathrm{psf}_{detector} = \begin{cases} \frac{a}{4\pi^2 x^2 + a^2} & : x <= 0 \\ 0 & : x > 0 \end{cases} \qquad (48)$$

$$\mathrm{psf}_{readout} = \begin{cases} a e^{-x/\beta} \sin(2\pi x/\gamma + \theta) & : x <= 0 \\ 0 & : x > 0 \end{cases} \qquad (49)$$

where

$$a = \frac{\beta\gamma(\gamma \sin(\theta) + 4\pi\beta\cos(\theta))}{\gamma^2 + 16\pi^2\beta^2} \qquad (50)$$

is the normalization factor, which ensures that the total integral of the $\mathrm{psf}_{readout}$ is equal to 1, k is the pixel value in real space, and a is the parameter of the Lorentzian function that describes the PSF of the detector. The parameters β, γ, and θ are the parameters of the damped harmonic oscillator, which is used to describe the PSF of the readout system. The model parameters were obtained by fitting to experimental images and by applying random variation to the fitting parameters.

Row-line noise. Row-line (RL) noise arises due to the non-response of the detector over some pixels during the scanning process along the fast-scan direction. This noise can be modeled by generating a random number of row lines with random length. The pixel intensities of the lines in the image are replaced by their average intensity multiplied by a random factor within the range [0.5, 1.5]. This can be represented as:

$$x \leftarrow \mathbb{RL}(x) \qquad (51)$$

Black pixel noise. Black pixels are randomly occurring pixels that have significantly lower values than their neighboring pixels,

causing black spots to appear in the image. These black pixels may result from information loss during data transmission, cosmic rays, or the detector's non-response. As black pixels are time-dependent, they can be modeled by generating a sensitivity mask ($S_{Blacknoise}$) with a spatially uniform distribution of a specified number of black points. This can be represented mathematically as:

$$x \leftarrow x S_{Blacknoise} \qquad (52)$$

However, in the case of SEM images, black spots in the images may be attributed to pores present in the sample, and hence, this type of distortion is not generated.

Zinger noise. Zingers are random white dots that appear in an image. They are caused by bursts of photons produced by electron-generated X-rays, cosmic rays, and muons in the scintillator[80]. Zinger noise can be simulated by creating a sensitivity mask ($S_{Zingernoise}$) with a spatially uniform distribution of a specified number of Zinger points.

$$x \leftarrow x S_{Zingernoise} \qquad (53)$$

Upper-clip noise. Upper-clip noise, also known as saturation noise, occurs when the intensity value of a pixel exceeds the maximum value that the analog-to-digital converter can detect. This causes the pixel to be clipped at the maximum value, resulting in an overly bright image with lost details. This type of noise can be modeled by setting a threshold value for the maximum intensity and clipping any pixel values above that threshold T_u.

$$x \leftarrow \min(x, T_u) \qquad (54)$$

Quantization noise. To generate an image in digital form, the analog voltage signal readout during the last stage is quantized into discrete values using an ADC with a gain a. This process introduces quantization noise.

$$x \leftarrow \mathrm{round}(ax) \qquad (55)$$

Figure 13 shows simulated STEM images of the different types of noise that can be found in STEM images. These distortions were

randomly added to the images to simulate real STEM conditions and make it easier to identify the different types of noise.

Post-processing distortions. Post-processing distortions are typically added after the image is recorded. These distortions, such as interpolation and blurring, can affect the noise in the image in a nonlinear way. Post-processing distortions can also include annotations and cropping, which replace part of the original image. Ideally, these distortions should be preserved by the restoration process. Interpolation distortions may happen when a user applies a transformation function to the image before it is restored. This might be done to make the image suitable for further post-processing or to better visualize an area of interest. Interpolation distortion can be modeled by applying a random transformation, such as a random linear transformation matrix, to the training image pair. Gaussian blurring is a way of distorting an image to reduce noise and improve the SNR. This is done by applying a 2D Gaussian function to the image with a given standard deviation σ. Although this type of blurring can improve the quality of an image, it can also alter the distribution of noise in the image. Therefore, when restoring an image, the blurring must be removed along with the distortion. In our training set, we only applied random σ values between 0 and 1 pixel to the distorted images. Annotations are added to an image to provide additional information or to highlight specific areas of the image. These can include text, shapes, and arrows, and may be added by the software or by the user. When creating training image pairs, we model the annotations by adding the same random annotations at the same pixel location in both the ground-truth and distorted images. Cropping is a type of post-processing distortion that involves removing one or more areas of an image. This can be done manually by the user or automatically in a processing workflow, such as after the image has been shifted, rotated or aligned. The removed areas are usually filled in with a constant value or the median of the image's value range. When creating training image pairs, we model this process by randomly replacing the intensity value in a randomly selected area in both images. The selected area is typically outside a central square or rectangle, such as 50% of the total image area, to mimic the fact that cropping is typically not applied to the central region, which may already be adjusted to show the main feature of interest.

DATA AVAILABILITY

The trained models for the six distinct neural networks described in this study are openly accessible. These models, together with illustrative scripts demonstrating their utilization, are provided to facilitate replication and further exploration. In addition, the script used for training these models is also included. All these resources are hosted in a dedicated GitHub repository, ensuring easy accessibility and version control. The repository can be accessed at the following URL: https://github.com/Ivanlh20/tk_r_em.

REFERENCES

1. Nellist, P. D. et al. Direct sub-angstrom imaging of a crystal lattice. *Science* **305**, 1741 (2004).
2. Joy, D. C. The aberration corrected SEM. In *AIP Conference Proceedings*, vol. 788, 535–542 (2005).
3. Kisielowski, C. et al. Detection of single atoms and buried defects in three dimensions by aberration-corrected electron microscope with 0. 5Å-information limit. *Microsc. Microanal.* **14**, 469–477 (2008).
4. Muller, D. A. et al. Room design for high-performance electron microscopy. *Ultramicroscopy* **106**, 1033–1040 (2006).
5. Jones, L. et al. Optimising multi-frame ADF-STEM for high-precision atomic-resolution strain mapping. *Ultramicroscopy* **179**, 57–62 (2017).
6. Zuo, J. M. Electron detection characteristics of a slow-scan CCD camera, imaging plates and film, and electron image restoration. *Microsc. Res. Tech.* **49**, 245–268 (2000).
7. Faruqi, A. R. & Subramaniam, S. CCD detectors in high-resolution biological electron microscopy. *Q. Rev. Biophys.* **33**, 1–27 (2000).
8. Ishikawa, R., Lupini, A. R., Findlay, S. D. & Pennycook, S. J. Quantitative annular dark field electron microscopy using single electron signals. *Microsc. Microanal.* **20**, 99–110 (2014).
9. Berkels, B. et al. Optimized imaging using non-rigid registration. *Ultramicroscopy* **138**, 46–56 (2014).
10. Jones, L. et al. Smart align-a new tool for robust non-rigid registration of scanning microscope data. *Adv. Struct. Chem. Imaging* **1**, 8 (2015).
11. Jones, L. & Nellist, P. D. Identifying and correcting scan noise and drift in the scanning transmission electron microscope. *Microsc. Microanal.* **19**, 1050–1060 (2013).
12. Braidy, N., Bouar, Y. L., Lazar, S. & Ricolleau, C. Correcting scanning instabilities from images of periodic structures. *Ultramicroscopy* **118**, 67–76 (2012).
13. Buban, J. P., Ramasse, Q., Gipson, B., Browning, N. D. & Stahlberg, H. High-resolution low-dose scanning transmission electron microscopy. *J. Electron Microsc.* **59**, 103–112 (2010).
14. Jong, A. D., Coene, W. & Dyck, D. V. Image processing of HRTEM images with non-periodic features. *Ultramicroscopy* **27**, 53–65 (1989).
15. Marks, L. D. Wiener-filter enhancement of noisy hrem images. *Ultramicroscopy* **62**, 43–52 (1996).
16. Kushwaha, H. S., Tanwar, S., Rathore, K. S. & Srivastava, S. De-noising filters for TEM (transmission electron microscopy) image of nanomaterials. In *Proceedings of the 2012 Second International Conference on Advanced Computing & Communication Technologies (ACCT)* (2012).
17. Du, H. A nonlinear filtering algorithm for denoising HR(S)TEM micrographs. *Ultramicroscopy* **151**, 62–67 (2015).
18. Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**, 2080–2095 (2007).
19. Mevenkamp, N. et al. Poisson noise removal from high-resolution STEM images based on periodic block matching. *Adv. Struct. Chem. Imaging* **1**, 1–19 (2015).
20. Kimoto, K. et al. Local crystal structure analysis with several picometer precision using scanning transmission electron microscopy. *Ultramicroscopy* **110**, 778–782 (2010).
21. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *CVPR*, 2261–2269 (IEEE Computer Society, 2017).
22. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, 833–851 (Springer Verlag, 2018).
23. Kim, D. W., Chung, J. R. & Jung, S. W. GRDN:grouped residual dense network for real image denoising and GAN-based real-world noise modeling. In *CVPRW*, 2086-2094 (IEEE Computer Society, 2019).
24. Cheng, Z., You, S., Ila, V. & Li, H. Semantic single-image dehazing. Preprint at http://arxiv.org/abs/1804.05624 (2018).
25. Luo, Z., Chen, S. & Qian, Y. A deep optimization approach for image deconvolution. Preprint at http://arxiv.org/abs/1904.07516 (2019).
26. Wang, X. et al. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
27. Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *International Conference on Machine Learning*, 807–814 (2010).
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach. Learn Res.* **15**, 1929–1958 (2014).
29. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, 448-456 (2015).
30. Wang, F. et al. Multi-resolution convolutional neural networks for inverse problems. *Sci. Rep.* **10**, 1–11 (2020).
31. Wang, F., Henninen, T. R., Keller, D. & Erni, R. Noise2atom: unsupervised denoising for scanning transmission electron microscopy images. *Appl. Microsc.* **50**, 23 (2020).
32. Mohan, S. et al. Deep denoising for scientific discovery: a case study in electron microscopy. *IEEE Trans. Comput. Imaging* **8**, 585–597 (2022).
33. Lin, R., Zhang, R., Wang, C., Yang, X.-Q. & Xin, H. L. TEMImageNet training library and AtomSegNet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images. *Sci. Rep.* **11**, 5386 (2021).

34. Gambini, L., Mullarkey, T., Jones, L. & Sanvito, S. Machine-learning approach for quantified resolvability enhancement of low-dose STEM data. *Mach. Learn. Sci. Technol.* **4**, 015025 (2023).

35. Kalinin, S. V. et al. Machine learning in scanning transmission electron microscopy. *Nat. Rev. Methods Primers* **2**, 1–28 (2022).

36. Mirza, M. & Osindero, S. Conditional generative adversarial nets. Preprint at http://arxiv.org/abs/1411.1784 (2014).

37. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization Preprint at https://arxiv.org/abs/1412.6980 (2014).

38. Vaswani, A. et al. Attention is all you need. In Guyon, I. et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).

39. Popel, M. & Bojar, O. Training tips for the transformer model. *Prague Bull. Math. Linguist.* **110**, 43–70 (2018).

40. Goodfellow, I. J. et al. Generative adversarial networks. In *Advances in neural information processing systems*, vol. 27 (2014).

41. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, Vol. 9351 LNCS, 234-241 (Springer International Publishing Switzerland, 2015).

42. Van Aert, S. et al. Quantitative atomic resolution mapping using high-angle annular dark field scanning transmission electron microscopy. *Ultramicroscopy* **109**, 1236–1244 (2009).

43. Martinez, G. T., De Backer, A., Rosenauer, A., Verbeeck, J. & Van Aert, S. The effect of probe inaccuracies on the quantitative model-based analysis of high angle annular dark field scanning transmission electron microscopy images. *Micron* **63**, 57–63 (2014).

44. De Backer, A., van den Bos, K. H., Van den Broek, W., Sijbers, J. & Van Aert, S. StatSTEM: an efficient approach for accurate and precise model-based quantification of atomic resolution electron microscopy images. *Ultramicroscopy* **171**, 104–116 (2016).

45. Krivanek, O. L. et al. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* **464**, 571–574 (2010).

46. Yamashita, S. et al. Atomic number dependence of *Z* contrast in scanning transmission electron microscopy. *Sci. Rep.* **8**, 1–7 (2018).

47. Lobato, I. & Van Dyck, D. MULTEM: a new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy* **156**, 9–17 (2015).

48. Lobato, I., van Aert, S. & Verbeeck, J. Progress and new advances in simulating electron microscopy datasets using MULTEM. *Ultramicroscopy* **168**, 17–27 (2016).

49. Hull, D. & Bacon, D. J. *Introduction to Dislocations*, 5th edn (Butterworth-Heinemann, 2011).

50. Hirel, P. Atomsk: a tool for manipulating and converting atomic data files. *Comput. Phys. Commun.* **197**, 212–219 (2015).

51. Van Aert, S. et al. Procedure to count atoms with trustworthy single-atom sensitivity. *Phys. Rev. B Condens. Matter Mater. Phys.* **87**, 064107 (2013).

52. Arslan Irmak, E., Liu, P., Bals, S. & Van Aert, S. 3D atomic structure of supported metallic nanoparticles estimated from 2D ADF STEM images: a combination of atom-counting and a local minima search algorithm. *Small Methods* **5**, 2101150 (2021).

53. De Backer, A. et al. Experimental reconstructions of 3D atomic structures from electron microscopy images using a Bayesian genetic algorithm. *npj Comput. Mater.* **8**, 1–8 (2022).

54. Gulrajani, I. & Lopez-Paz, D. In search of lost domain generalization. *ICLR 2021 - 9th International Conference on Learning Representations* (2020).

55. Kalinin, S. V. et al. Automated and autonomous experiments in electron and scanning probe microscopy. *ACS Nano* **15**, 12604–12627 (2021).

56. Altantzis, T. et al. Three-dimensional quantification of the facet evolution of Pt nanoparticles in a variable gaseous environment. *Nano Lett.* **19**, 477–481 (2019).

57. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void – learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2129–2137 (2019).

58. Cayado, P. et al. Critical current density improvement in CSD-grown high-entropy REBa $_2$Cu$_3$O$_{7-\delta}$ films. *RSC Adv.* **12**, 28831–28842 (2022).

59. Grünewald, L. *Electron Microscopic Investigation of Superconducting Fe- and Cu-based Thin Films*. Ph.D. thesis, Laboratorium für Elektronenmikroskopie (LEM), Karlsruher Institut für Technologie (KIT) (2022).

60. Molina-Luna, L. et al. Atomic and electronic structures of BaHfO$_3$-doped TFA-MOD-derived YBa $_2$Cu$_3$O$_{7-\delta}$ thin films. *Supercond. Sci. Technol.* **28**, 115009 (2015).

61. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **3**, 47–57 (2016).

62. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976 (2017).

63. Zhang, Y., Tian, Y., Kong, Y., Zhong, B. & Fu, Y. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2472–2481 (2018).

64. Ignatov, A. & Timofte, R. Ntire 2019 challenge on image enhancement: methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2019).

65. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. CBAM: convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).

66. Plötz, T. & Roth, S. Neural nearest neighbors networks. In (eds Bengio, S. et al.) *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Associates, Inc., 2018).

67. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).

68. Li, Z., Fan, Y. & Liu, W. The effect of whitening transformation on pooling operations in convolutional autoencoders. *EURASIP J. Adv. Signal Process* **2015**, 1–11 (2015).

69. Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GaN. Preprint at https://arxiv.org/abs/1807.00734v3 (2018).

70. Jarrett, K., Kavukcuoglu, K., Ranzato, M. & LeCun, Y. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th International Conference on Computer Vision*, 2146–2153 (2009).

71. Van Aert, S., Batenburg, K. J., Rossell, M. D., Erni, R. & Van Tendeloo, G. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature* **470**, 374–377 (2011).

72. Beyer, A., Straubinger, R., Belz, J. & Volz, K. Local sample thickness determination via scanning transmission electron microscopy defocus series. *J. Microsc.* **262**, 171–177 (2016).

73. Lobato, I. & Van Dyck, D. An accurate parameterization for scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints. *Acta Crystallogr. A* **70**, 636–649 (2014).

74. Hartel, P., Rose, H. & Dinges, C. Conditions and reasons for incoherent imaging in STEM. *Ultramicroscopy* **63**, 93–114 (1996).

75. Chew, L. P. & Drysdale, R. L. Voronoi diagrams based or convex distance functions. *Proceedings of the 1st Annual Symposium on Computational Geometry, SCG 1985* 235–244 (1985).

76. Spence, J. C. H. & Zuo, J. M. Large dynamic range, parallel detection system for electron diffraction and imaging. *Rev. Sci. Instrum.* **59**, 2102–2105 (1988).

77. Tietz, H. Design and characterization of 64 megapixel fiber optic coupled cmos detector for transmission electron microscopy. *Microsc. Microanal.* **14**, 804–805 (2008).

78. Clough, R. N., Moldovan, G. & Kirkland, A. I. Direct detectors for electron microscopy. *J. Phys. Conf. Ser.* **522**, 012046 (2014).

79. Thust, A. High-resolution transmission electron microscopy on an absolute contrast scale. *Phys. Rev. Lett* **102**, 220801 (2009).

80. Vulovic, M., Rieger, B., Van Vliet, L. J., Koster, A. J. & Ravelli, R. B. G. A toolkit for the characterization of CCD cameras for transmission electron microscopy. *Acta Cryst. D* **66**, 97–109 (2010).

81. Kodak. CCD Image Sensor Noise Sources. *Application Note* (2005).

82. Konnik, M. & Welsh, J. High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial. Preprint at https://arxiv.org/pdf/1412.4031 (2014).

83. Gow, R. D. et al. A comprehensive tool for modeling CMOS image-sensor-noise performance. *IEEE Trans. Electron. Devices* **54**, 1321–1329 (2007).

84. Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. A model for measurement of noise in ccd digital-video cameras. *Meas. Sci. Technol.* **19**, 045207 (2008).

85. Zinger, A. Detection of best and outlying normal populations with known variances. *Biometrika* **48**, 457–461 (1961).

86. Yule, G. U. On a method of investigating periodicities in disturbed series. *Philos. Trans. Royal Soc. A* **226**, 167–298 (1927).

87. Walker, G. On periodicity in series of related terms. *Proc. Royal Soc. A* **131**, 518–32 (1931).

88. Industries, B. *Photomultiplier Handbook* 1st edn (Burle Technologies Inc., 1980).

89. Mittelberger, A., Kramberger, C. & Meyer, J. C. Software electron counting for low-dose scanning transmission electron microscopy. *Ultramicroscopy* **188**, 1–7 (2018).

90. Mullarkey, T., Downing, C. & Jones, L. Development of a practicable digital pulse read-out for dark-field STEM. *Microsc. Microanal.* **27**, 99–108 (2021).

91. Krivanek, O. L. et al. An electron microscope for the aberration-corrected era. *Ultramicroscopy* **108**, 179–195 (2008).

92. Ede, J. M. Warwick electron microscopy datasets. *Mach. Learn. Sci. Technol.* **1**, 045003 (2020).

93. Aversa, R., Modarres, M. H., Cozzini, S., Ciancio, R. & Chiusole, A. The first annotated set of scanning electron microscopy images for nanoscience. *Sci. Data* **5**, 1–10 (2018).

94. Amini, M. N. et al. Understanding the effect of iodide ions on the morphology of gold nanorods. *Part. Part. Syst. Charact.* **35**, 1800051 (2018).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS
I.L. and S.V.A. designed the study. I.L. created the mathematical models for the undistorted and distorted EM images, implemented, trained, and evaluated the NN models. T.F. conducted the quantitative analysis of STEM images for the models. All authors contributed to the planning and execution of the research, discussed the results, and helped in writing the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-023-01188-0.

Correspondence and requests for materials should be addressed to I. Lobato or S. Van Aert.

Reprints and permission information is available at http://www.nature.com/reprints

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