The Gold Standard for low-dose STEM

Akshaya Kumarjaishankar(akshaya.kumarjaishankar@ru.nl)

Willem de Kleijne(**w.p.m.dekleijne@tudelft.nl**)

Jay te Beest(**j.t.te.beest@liacs.leidenuniv.nl**)

Avital Wagner(avital.wagner@radboudumc.nl)

Motivation

Biological specimens are highly susceptible to electron-beam damage, limiting the use of scanning transmission electron microscopy (STEM). We target imaging at ultra-low dose (<1 $e^{-}/Å^{2}$) by reducing dwell time and emission current. However, short dwell times introduce severe scan artifacts and low signal-to-noise, even on a gold-standard sample (gold islands on amorphous carbon). Artifacts include: flyback pixels, bleed through and streaking, vertical line artifacts from low detector signal, geometric distortions (stretching), overall intensity shifts (lower median intensity), and reduced apparent resolution due to low signal-to-noise.

Here we describe a workflow that combines targeted Fourier-domain pre-processing with a supervised deep-learning model trained on real data to suppress scan artifacts while preserving texture and resolution.

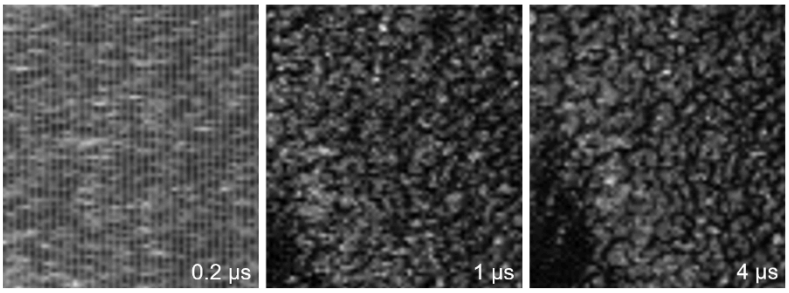


Figure 1. STEM images (2k) taken with a HAADF detector of gold islands on amorphous carbon using 200 kV and dwell times of 0.2, 1, and 4 µs which correspond to electron doses of 0.136, 0.681, and 2.723 $e^{-}/Å^{2}$.

STEM data set – Because simulating the full range of scan artifacts is challenging, we trained directly on real paired data. sample: gold islands on amorphous carbon; detector: HAADF. Dwell times: 0.2, 0.5, 1, 2, and 4 µs. Image sizes: 512×512, 1024×1024, and 2048×2048. Acceleration voltages: 200 kV.

Workflow and Methods

Pre-processing - we observed 1-pixel-wide vertical line artifacts repeating every other column, attributable to low HAADF detector signal. In Fourier space, these lines appear as isolated high-frequency components. We removed them using a simple frequency mask, masking last ~35% of spatial frequencies.

Training set – to increase the effective dataset size, we extracted patches (128x128 px) from the 2k x 2k px images. Since we have few images, we sample the training

data many times, each epoch sees 3 (train images) times 10k (repeats) = 30k pairs of patches. We train for 50 epochs due to time constraints.

Importantly, patches with an average intensity below some threshold, which is put just below the total patch average intensity, were omitted to avoid training on grid bars rather than features of interest.

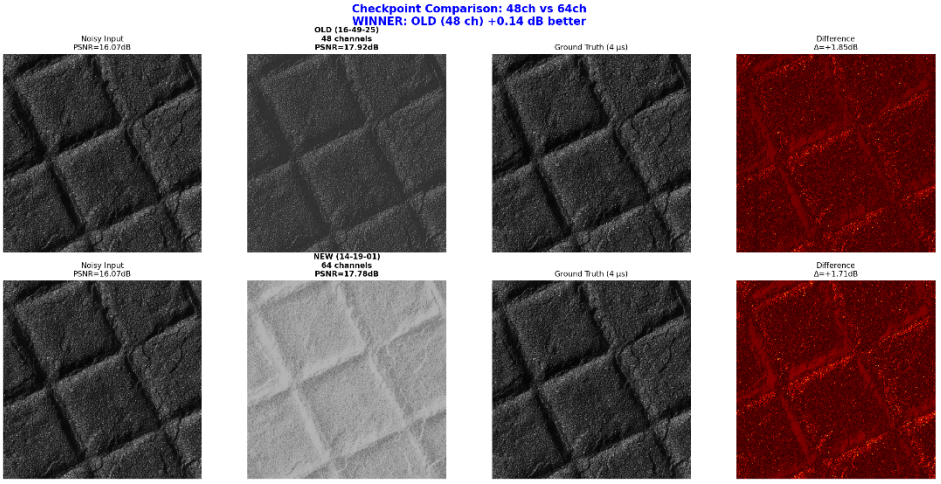
Model – we implemented a U-Net–type architecture following Noise2Noise (Lehtinen et al., 2018), with five encoding and decoding layers and skip connections. The network employs 48 base channels, 2×2 max pooling for downsampling, nearest-neighbour upsampling in the decoder, and LeakyReLU activations ($\alpha = 0.1$). All convolutional layers use 3×3 kernels.

Supervised learning setup – Inputs were low dwell time images (0.2, 0.5, and 2 µs) and the target was the corresponding 4 µs image. Images at 1 µs were held out for inference.

Loss function design-Standard pixel-wise losses (L1/L2/MSE) produced overly smooth images and removed fine texture. To better preserve high-frequency information, we designed a custom loss that is a weighted sum of two terms **(1)** a Fourier-domain similarity term with **(2)** an explicit resolution constraint. The former computes the L1 loss between the Fourier magnitudes of the predicted image and the target image. Using Fourier magnitude makes this term less sensitive to small spatial misalignments between images. The later we include a resolution term based on the radially averaged Fourier magnitude. The noise level is estimated from the high-frequency part of the spectrum, and the resolution cutoff is defined as the highest spatial frequency at which the radial profile exceeds three times this noise level. The loss penalizes reconstructions whose cutoff frequency is lower than that of the target.

Results

Model Selection After systematic evaluation of multiple training runs, checkpoint 2025-12-17_16-49-25 demonstrated optimal performance. Performance on held-out 1 µs test data: PSNR: 16.07 → 17.92 dB (+1.85 dB improvement) SSIM: 0.093 → 0.143 (+0.050) SNR: 3.40 → 5.25 dB (+1.85 dB) The model effectively suppresses scan artifacts while preserving structural details.



Outlook and Future Work

We further customized the loss function by adding a term that measures resolution anisotropy along the horizontal (x) and vertical (y) directions. Because the dominant artifact appears as smearing along x, this axis typically shows lower resolution than y. In clean images, the resolution is expected to be approximately isotropic. We therefore penalize differences between the estimated cutoff frequencies along x and y, encouraging comparable resolution in both directions. In addition to loss design, model performance could be improved by exploring alternative architectures, such as U-Net variants. Increasing the amount of training data would also likely improve robustness and generalization. Finally, additional data pre-processing, such as excluding featureless regions of the images, may help focus training on informative structures and reduce the influence of background noise.