

Unsupervised neural network-based image restoration framework for pattern fidelity improvement and robust metrology

Zijian Du^{✉,*}, Lingling Pu, Paul Wei[✉], Rui Yuan, Jeeeon Kim, and Jiaoying Tan
ASML-Silicon Valley, San Jose, California, United States

ABSTRACT. **Background:** Scanning electron microscope (SEM) images acquired by E-beam tools for inspection and metrology applications are usually degraded by blurring and additive noises. Blurring sources include the intrinsic point spread function of optics, lens aberration, and potential motion blur caused by the wafer stage movements during the image acquisition process. Noise sources include shot noise, quantization noise, and electronic read-out noise. Image degradation caused by blurring and noise usually leads to noisy, inaccurate metrology results. For low-dosage metrology applications, metrology algorithms often fail to obtain successful measurements due to elevated levels of blurring and noise. Image restoration and enhancement are necessary as preprocessing steps to obtain meaningful metrology results. Initial success was obtained by applying neural network-based framework to drastically improve image quality and metrology precision as is demonstrated in the previous work.

Aim: We aim to provide more details on the neural network model architecture, model regularization, and training dynamics to better understand the model's behavior. We also analyze the effect of image restoration on key metrology performances such as line edge roughness and mean critical dimension of the patterns.

Approach: Non-machine learning-based image quality enhancement methods fail to restore low-quality SEM images to a satisfactory degree. More recent convolutional neural networks and vision transformer-based, supervised deep learning models have achieved superior performance in various low-level image processing and computer vision tasks. Nevertheless, they require a huge amount of training data that contain high-quality ground truth images. Unfortunately, high-quality ground truth images for low-dosage SEM images do not exist. Instead, we use self-supervised U-Net combined with a fully connected network (FCN) to recover low-dosage images without the need for ground truth training images. The methodology can be applied to various one- and two-dimensional patterns with different scales, shapes, spatial density, and image intensity statistics. We use image quality metrics and loss function to guide model architecture optimization and study how regularization strength affects the restoration process. These studies provide a better understanding of how the model learns to restore images and how parameters and hyperparameters affect results.

Results: It is demonstrated that image quality metrics could be successfully used to evaluate self-supervised image restoration process and determine stopping criteria. The restored images show significantly improved image quality and metrology performance. Together, these pave the road to a systematic and automatic implementation of this methodology in real metrology applications.

Conclusions: A self-supervised U-Net-based model combined with FCN proved itself as a powerful tool to restore highly blurry and noisy low-dosage SEM images. It can be used to improve image quality, suppress metrology noise, and provide

*Address all correspondence to Zijian Du, zijian.du@asml.com

more robust measurements. It would become a crucial and necessary preprocessing step for metrology tasks as the E-beam dosage decreases and image quality worsens.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMM.22.3.034201](https://doi.org/10.1117/1.JMM.22.3.034201)]

Key words: deep learning; U-Net; convolutional neural networks; fully connected network; image restoration; deblurring; self-supervised model; generative modeling; metrology

Paper 22028G received Jun. 22, 2022; revised May 31, 2023; accepted Aug. 1, 2023; published Aug. 21, 2023.

1 Introduction

The progressive scaling of transistors in semiconductor manufacturing pushes the limit of lithographic techniques capable of smaller feature patterning with feature sizes well below 7-nm nodes. Electron-beam imaging has been established as the technology of choice for in-line and off-line metrology tasks. Highly massive metrology results are necessary to evaluate the process quality of printed resist patterns. With the scaling of semiconductor manufacturing technology, requirements for metrology and tolerance for edge placement error (EPE) are becoming increasingly demanding. High accuracy, precision, repeatability, and fast turn-around time are desirable. As the patterns scale, a lower electron dosage is needed to avoid photoresist damage. Under such conditions, the scanning electron microscope (SEM) images contain an excessive amount of noise and blurring. Electronic, thermal, mechanical, and quantization noise all contribute to images with low signal-to-noise ratio (SNR). Furthermore, E-beam aberration as well as relative movement between the wafer stage and detector introduces optical and motion blur to the images and causes high uncertainty for the edge profile of printed resist patterns. These pose challenges to SEM image pre-processing algorithms for high-quality metrology tasks. In extreme conditions, the SEM images are simply unmeasurable with edge profile-based metrology algorithms. Hence, an effective image restoration methodology is crucial to turning unmeasurable SEM images metrology-ready and effectively improving metrology performance to meet demanding specifications for massive metrology. Below, we give a high-level review of existing technology for image quality (IQ) enhancement and restoration.

1.1 Non-Deep Learning Based Image Restoration

Before machine learning (ML) became the dominant methodology for every field in science and technology, the problem of image restoration has already been extensively studied. Image deblurring methodology can be classified into non-blind deblurring and blind deblurring using non-deep learning (DL)-based architecture. The first category describes a flow that makes explicit assumptions on the function of the underlying blurring kernel and uses iterative optimization or Fourier domain inverse filtering to solve for restored image.^{1,2} The second category does not explicitly assume a function for the kernel and usually describes a flow using a Bayesian approach. Usually, complex optimization algorithms are needed to solve for the solution.³ Both approaches are considered ill-posed inverse problems.

Most image restoration research models the image degradation process as a convolution of the underlying sharp image with degradation kernel plus additive noise term introduced by the SEM image acquisition process. As is shown in Eq. (1), x represents latent sharp and clean image, k models the kernel, η models noise intrinsic to the SEM image, and y is the captured SEM image. In a signal restoration process, the goal is to recover the underlying image x . Figure 1 shows the described image degradation process Eq. (1):

$$y = k \otimes x + \eta. \quad (1)$$

Usually this is an ill-posed, highly under-constrained problem hence it is necessary to reduce the solution space to obtain meaningful solutions. Assumptions are usually made on noise form, degradation kernel, and sometimes on image itself. Another approach is to solve the maximum a posterior problem as in Eq. (2), in which the latent image has a conditional probability

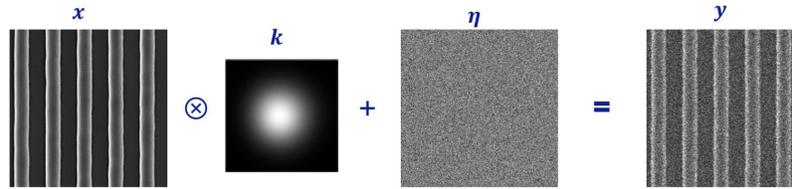


Fig. 1 Illustration of the image degradation process.

distribution density $P(x|y, k)$ given assumptions on kernel and degraded image. The proper choice of prior term $P(x)$ has been researched extensively to ensure the correct model. Among many proposed terms, the most widely used is l_p norm and its variants. This is based on the observation that gradient intensity of natural image follows a Laplace distribution as in Eq. (3).⁴

$$P(x|y, k) = \frac{P(y|x, k)P(x)}{P(y)}, \quad (2)$$

$$p^{\text{Lap}}(\nabla x) = \prod_i \frac{1}{2b} \exp\left(-\frac{\|\nabla x_i\|_1}{b}\right). \quad (3)$$

To recover the sharp image, many non-blind deblurring approaches try to solve the optimization problem with objective functions in the form of Eq. (4). Total variation (TV) prior of sparse image gradient as well as l_2 norm prior on image pixel intensity are popular choices.^{2,5,6} An improved regularization terms l_1/l_2 are used instead of l_1 or l_2 alone which both tend to over-regularize gradient while l_1 scaled by l_2 term has a better chance of converging to sharp restored image instead of smooth and blurry solution.⁴

$$\hat{x} = \arg \max_x \log p(y|x) + \log p(x). \quad (4)$$

All the above-mentioned methodologies assume an initial kernel form then iteratively optimizes the kernel to converge to underlying true kernel using appropriate image prior as regularization. More recent research has been focusing on solving blind restoration task where kernel form is not assumed.^{7,8} This is an even more challenging task since we want to model both kernel and latent image distribution given degraded image only. From a Bayesian perspective, a generative model is needed to model joint distribution of both kernel and latent image. It was pointed out that to successfully model $P(x, k|y)$ as in Eq. (5), we can never collect enough measurements because the number of unknowns grows exponentially with image size, hence instead of modeling $\text{MAP}_{k,x}$, we can model MAP_k which has much lower dimension and fewer unknown since kernel size is much smaller than image itself.⁹ After solving for MAP_k , x can then be solved by Fourier technique. In Ref. 10, instead of modeling image prior, the author modeled image gradients. It has been observed that image gradient obeys heavy-tailed distribution with most of its masses concentrated on small values but significantly higher probability at large values than Gaussian distribution, due to abrupt image intensity variations caused by strong image features such as edges and corners. The paper tried to solve restoration problem from a Bayesian approach and modeled latent image gradient as mixture of Gaussians and assumed exponential distribution for the kernel to induce sparsity and smoothness.¹⁰ As is shown in Eq. (6), the first term on the right-hand side is Gaussian term on gradient of degraded image, second term is mixture of Gaussians on restored image gradient, and third term is exponential function assumption on kernel. The authors tried to approximate the full posterior distribution $P(K, \nabla X | \nabla Y)$ then compute kernel with maximum marginal probability:

$$P(x, k|y) = \frac{P(y|x, k)P(x)P(k)}{P(y)}, \quad (5)$$

$$P(K, \nabla X | \nabla Y) \propto P(\nabla Y | K, \nabla X) P(\nabla X) P(K) = \prod_i N(\nabla Y(i) | (K \otimes \nabla X(i)), \sigma^2) \times \prod_i \sum_{c=1}^C \pi_c N(\nabla X(i) | 0, \nu_c) \times \prod_j \sum_{d=1}^D \pi_d E(K_j | \lambda_d). \quad (6)$$

In real situations, blurring kernel can be arbitrary without specific form and this is especially true for random motion blur. Hence non-blind deblurring usually suffers in performance due to simplistic or simply wrong assumption of kernel. Another challenge is to apply optimization algorithm with appropriate regularization strength to converge to the desired solution. On the other hand, blind deblurring loosens strong prior term but requires more sophisticated variational inference algorithm. The two solutions contain intrinsic trade-off of over and under regularization. Next section will go through some typical optimization techniques used in non-blind deblurring tasks.

1.2 Iterative Optimization for Non-Blind Deblurring

Non-blind deblurring usually resorts to optimization algorithms such as half quadratic splitting (HQS) or alternating direction method of multipliers.^{3,5} A routine practice is to initialize degradation kernel as 2D Gaussian function and degraded image as initial latent image then iteratively solve for kernel and latent image alternatively while fixing the other term. Authors in Ref. 11 took advantage of the denoising effect of U-Net and used diluted convolutional layers to increase receptive fields of the convolutional neural networks (CNNs).¹² Such an optimization process can be formulated as Eqs. (7) and (8). Both methods optimize data fidelity subproblem in HQS using CNNs, as in Eq. 9(a) and solve for sharp image using inverse filtering. Equation 9(b) corresponds to Gaussian denoising on x_k with noise level of $\sqrt{\lambda/\mu}$, hence any CNN inspired Gaussian denoiser can be plugged into the formula to solve Eq. 9(b). This methodology has been extended to solve other image restoration problems such as de-mosaicking and image super-resolution.¹¹ Prior regularization term is routinely added to optimization formulation to reduce solution choices and stabilize convergence to avoid noise amplification or trivial solution which are common results for its unregularized counterpart:

$$\hat{x} = \arg \min_x \frac{1}{2\sigma^2} \|y - k \otimes x\|^2 + \lambda R(z) \quad \text{s.t. } z = x, \quad (7)$$

$$L_\mu(x, z) = \frac{1}{2\sigma^2} \|y - k \otimes x\|^2 + \lambda R(z) + \frac{\mu}{2} \|z - x\|^2, \quad (8)$$

$$\begin{cases} x_k = \arg \min_x \|y - k \otimes x\|^2 + \mu\sigma^2 \cdot \|x - z_{-1}\|^2, & (a) \\ z_k = \arg \min_z \frac{1}{2(\sqrt{\lambda/\mu})^2} \|z - x_k\|^2 + R(z). & (b) \end{cases} \quad (9)$$

1.3 Deep Learning Based Image Restoration

Due to high heterogeneity of image statistics, a fixed prior form usually cannot model true image prior, and this leads to wrong model and unsatisfactory results. This problem led researchers to experiment with a data-driven approach to model prior term. In Refs. 6 and 13, the author used CNNs to model the prior of training images. In Ref. 14, the author generalized shrinkage fields by removing unnecessary parameter sharing and replacing pixel-wise applied shrinkage functions with CNN that applied to entire image. These approaches outperform the results obtained from hand-crafted prior term approaches, illustrating flexible modeling capability of CNNs.

To further utilize modeling capability of CNNs, end-to-end image restoration frameworks have been proposed and researched. In this framework, CNN is treated as black box and takes low quality image as input while outputs restored high quality images. Convolutional layer followed by element-wise rectified linear unit (ReLU) layer, sum of data fidelity term, and l_2 regularization as loss function was used.^{15,16} Researchers used powerful generative adversarial network to restore image by optimizing adversarial information loss.¹⁷ Reference 18 used CNN as feature extraction module that extracts image features from degraded images then estimates kernel and latent image. Reference 19 used six-layers CNNs to learn gradient map and obtained kernel using Fourier transform and HQS to obtain restored image. CNN empowered approaches,

when properly combined with physical model, can achieve state-of-the-art results compared with methods that assume prior explicitly and can provide end-to-end solutions.^{20,21} However, the supervised method carries crucial pitfalls that make it undesirable for SEM image restoration task.

The trained model suffers from poor generalization to images and kernel forms it was not trained on. This limits the real-world applications of these approaches since realistic blurring kernels in SEM images could be arbitrary without fixed patterns.^{22,23} Another crucial challenge in restoring low quality SEM images is the lack of ground truth images and supervised ML model does not work without ground truth. Therefore, it becomes obvious that a pre-training free, self-supervised, generative methodology that imposes just enough prior assumption is strongly preferred in SEM image restoration applications. In this paper, we present a thorough investigation of such a SEM image restoration methodology and detailed evaluation of its performance.

2 Dataset and Methodology

In this section, we describe in detail (1) SEM image dataset, (2) architecture of the neural networks, (3) IQ evaluation metrics, and (4) model regularization and convergence.

2.1 SEM Image Dataset

SEM images were collected from 10 dies and 10 runs for each die and four images were captured for each run, resulting in 400 images. The images were captured using ASML HMI eP5 metrology and inspection machine. The patterns include AEI and ADI line space (LS) and contact holes (CH). Pixel size is 1 nm, the field of view is 2048×2048 , there are 19 line patterns in each LS image and 1024 CHs in each CH image. To create a dataset with improving image qualities, we used frame averaging to obtain 1, 2, 3, and 4 frame averaged images from the original datasets. Figure 2 summarizes the entire dataset.

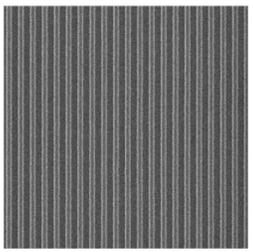
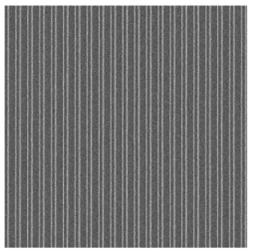
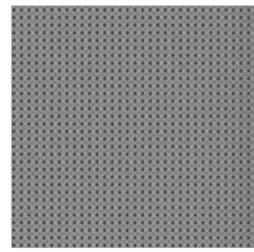
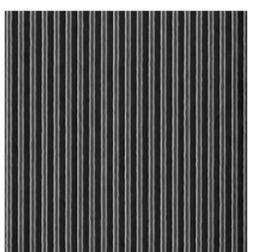
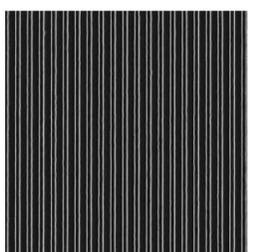
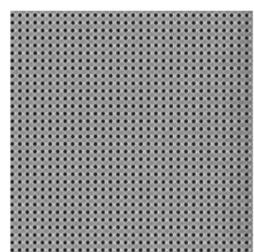
Pattern	<i>LS AEI</i>	<i>LS ADI</i>	<i>CH</i>
Pixel Size	1 nm	1 nm	1 nm
FoV	2048×2048	2048×2048	2048×2048
Low dose images			
Restored images			

Fig. 2 Low dosage SEM image dataset.

Table 1 Architecture of generative network G_k .

Input: kernel size $k \times k$, z_k (200) from the uniform distribution with fixed seed
Output: blur kernel k with size $k \times k$
Hidden layer: linear (200, 1000); ReLU
Output layer: linear (1000, $k \times k$); SoftMax
Reshape 1D output to obtain 2D blur kernel with size $k \times k$

2.2 Neural Network Architecture

The proposed image restoration framework is composed of generative networks G_k which generates kernel and G_x which generates latent image.

Generative network G_k is used to generate blurring kernel k , which has much lower dimension than latent image hence should be modeled using a lightweight fully connected network (FCN). The FCN takes a one-dimensional (1D) noise vector z_k with 200 dimensions as input and a hidden layer of 1000 nodes and an output layer of k^2 nodes. To guarantee the non-negativity constraint, the SoftMax nonlinearity is applied to the output layer of G_k . Finally, the 1D output of k^2 entries is reshaped to a 2D $k \times k$ blurring kernel. Table 1 shows the architecture details.

Generative network G_x is an asymmetric autoencoder with skip connections and is used to generate latent clean image.²⁴ The first five layers of encoder are skip-connected to the last five layers of decoder. Finally, a convolutional output layer is used to generate a latent clean image. Since the output image needs to have positive values, the Sigmoid nonlinearity is applied to the output layer. As is shown in Fig. 3, the i 'th unit encoder-decoder architecture is illustrated. Taking e_i as an example, we use the form $e_i(n_f, k, p)$ to represent that the convolution in e_i has n_f filters with kernel size $k \times k$ and $p \times p$ padding. The filter size in the last convolutional layer is fixed as 1×1 since we apply Sigmoid activation function to each pixel without any spatial averaging to avoid blurring the image. Down-sampling and up-sampling both have a stride of 2 and bilinear interpolation is used for up-sampling. The input dimension is c_{in} and the filter number at i 'th skip connection layer is proportional to the filter number at i 'th encoder layer with ratio α . This ratio should determine the proportion of the information i 'th decoder layer obtains directly from i 'th encoder layer. Kernel size is fixed at 3×3 and padding fixed at 1. Neural network architecture usually has major impacts on the results and in this case the restored image's quality. We picked key architectural hyperparameters that include input dimension c_{in} , filter number combination for each layer (n_1, n_2, n_3, n_4, n_5), and the ratio of the filter number between skip layer and encoder layer α . We experimented with a series of choices and chose the best combination guided by IQ metrics. Since SEM images are single channel gray level images, the input has single channel. The architecture details are presented in Table 2, and the illustration is shown in Fig. 4. The choice of model architecture is determined by factors, such as pattern complexity, computational latency, and restoration performance. For 1D patterns, $c_{in} = 8$ and filter number is (4, 4, 8, 16, 16), while for 2D patterns, $c_{in} = 32$ and filter number is (8, 8, 16, 32, 32). We also set $\alpha = 0.5$ since it provides the best restored image while still maintaining enough pattern style from the low-dosage image.

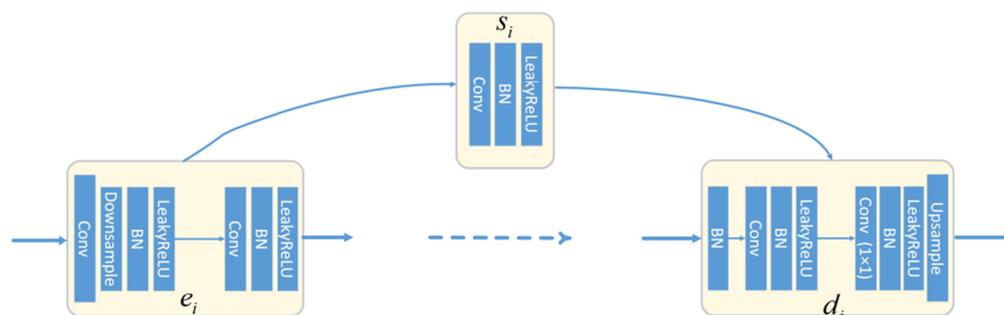


Fig. 3 Schematic illustration of encoder, skip, and decoder layers.

Table 2 Architecture of generative network G_x .

Input: z_x ($c_{in} \times 2048 \times 2048$) From the uniform distribution with fixed seed.

Output: latent image x ($1 \times 2048 \times 2048$) modeled by the neural network

Encoder unit 1 $e_1(n_1, 3, 1), s_1(\alpha n_1, 1, 1)$

Encoder unit 2 $e_2(n_2, 3, 1), s_2(\alpha n_2, 1, 1)$

Encoder unit 3 $e_3(n_3, 3, 1), s_3(\alpha n_3, 1, 1)$

Encoder unit 4 $e_4(n_4, 3, 1), s_4(\alpha n_4, 1, 1)$

Encoder unit 5 $e_5(n_5, 3, 1), s_5(\alpha n_5, 1, 1)$

Decoder unit 5 $d_5(n_5, 3, 1)$

Decoder unit 4 $d_4(n_4, 3, 1)$

Decoder unit 3 $d_3(n_3, 3, 1)$

Decoder unit 2 $d_2(n_2, 3, 1)$

Decoder unit 1 $d_1(n_1, 3, 1)$

Output layer Conv. ($n_1, 1, 1$); Sigmoid

Input dimension c_{in} {4, 8, 32}

Filter number $((n_1, n_2, n_3, n_4, n_5))$ (4, 4, 8, 16, 16), (8, 8, 16, 32, 32)

Filter number ratio α {0.25, 0.5, 0.75}

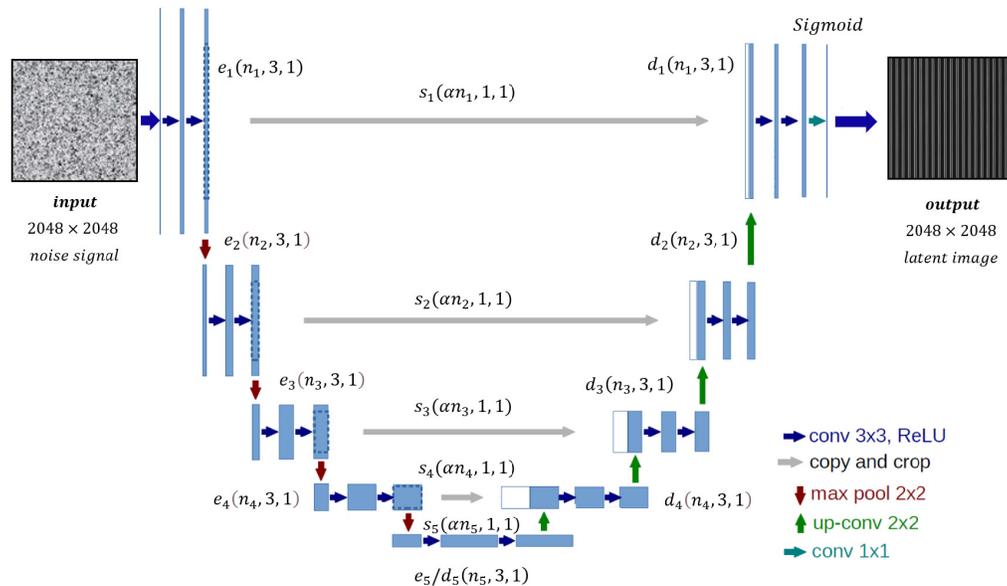

Fig. 4 Architecture of G_x .

Image restoration network is composed of G_x and G_k , the latent image z_x convolves with kernel z_k , resulting in a blurred image, and we use its mean squared error (MSE) with low-quality SEM image y as the image fidelity term. We formulate the image restoration into an unconstrained optimization problem. It can be seen from the introduction that appropriate regularization strength is crucial to obtaining satisfactory results, and we add TV regularization terms for both image and kernel and weigh the two terms by coefficients λ_x and λ_k , respectively, to encourage sparsity. The choice of λ_x should be related to the noise level in image y with larger λ_x for a higher noise level to prevent overfitting the noise in the image. The choice of λ_k should reflect the

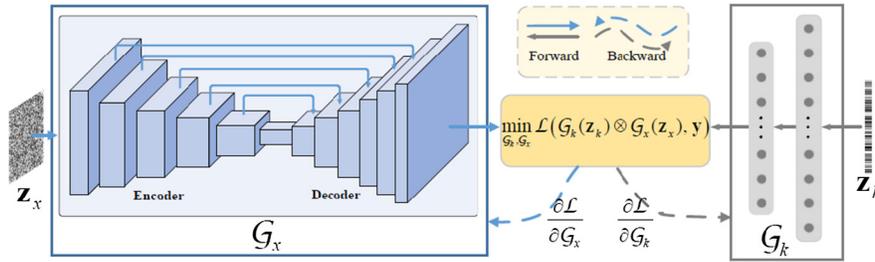


Fig. 5 Schematic illustration of the two generative networks and training process.

Table 3 Pseudo code for joint optimization.

Input: Low dosage SEM image y

Output: Blur kernel k and clean image x

- 1: Sample z_k and z_x from uniform distribution with fixed seed
- 2: **for** $t = 1$ to T **do**
- 3: $k = G_k^{t-1}(z_k)$
- 4: $x = G_x^{t-1}(z_x)$
- 5: Compute the gradients w.r.t. G_k and G_x
- 6: Update G_k^t and G_x^t using ADAM algorithm
- 7: **end for**
- 8: $x = G_x^T(z_x)$, $k = G_k^T(z_k)$

complexity of the kernel and since the kernel is supposed to be simple for SEM image, the regularization term could be fixed at a higher value. Our image restoration loss function is composed of the MSE term in addition to two TV regularization terms; the goal is to minimize the total loss, which can be written as Eq. (10):

$$\min_{G_k, G_x} \|G_k(z_k) \otimes G_x(z_x) - y\|^2 + \lambda_x \text{TV}(G_x(z_x)) + \lambda_k \text{TV}(G_k(z_k)). \quad (10)$$

The optimization process for Eq. (10) can be viewed as a “zero-shot,” self-supervised learning,²⁵ where both the generative networks are trained using only the low-quality SEM image and no ground-truth clean image. To optimize the networks, we adopt joint optimization which takes advantages of the automatic differentiation technique, the gradients w.r.t. G_k and G_x can be derived and the network parameters updated. The reason alternating minimization is not used is because this is an unconstrained problem, and it is easy to get stuck at saddle point due to the highly non-convex nature of loss function. We picked the ADAM optimization algorithm to simultaneously update G_k and G_x in one step due to its gradient and learning rate adaptive nature by using momentum.²⁶ Table 3 shows the pseudo code for the optimization process. Both the generative networks and the training process are schematically illustrated in Fig. 5.

2.3 Image Quality Evaluation Metrics

We describe several popular IQ evaluation metrics that will be used to evaluate the image restoration process and guide the learning process.

2.3.1 Peak signal-to-noise ratio

Peak signal-to-noise ratio (PSNR) is calculated by Eq. (11), m and n are the image dimensions, and MAX is the maximum signal value, which is 255 for grayscale 8 bit SEM images. X and Y represent the pixel value matrix of the reference image and the image to be measured,

respectively. High PSNR value indicates a low noise level in the measured image compared to the reference image. We picked this metric since we want to denoise low-dosage SEM images and hope to achieve high PSNR:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right), \quad \text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i, j) - Y(i, j)]^2. \quad (11)$$

2.3.2 Structural similarity index measure

Structural similarity index measure (SSIM) is calculated by Eq. (12), $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ are mean and variance of image matrix X and Y , σ_{xy}^2 is the covariance term of the two, and c_1 and c_2 are two scalars to stabilize division numerically. SSIM measures the relative geometrical information similarity between test image w.r.t. reference image and the value ranges from 0 to 1. We picked this metric since we want to restore the underlying patterns from the low-dosage SEM image with high integrity. Since SSIM is sensitive to pattern shift, this metric helps to identify potential pattern misalignment which could cause issues in metrology:

$$\text{SSIM}(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (12)$$

2.3.3 Pattern sharpness

Pattern sharpness (PS) is a measure to evaluate quality of the pattern edges. Figure 6 shows how this measure is calculated. It is calculated based on the gray-level profile extracted at pattern edge, and the value is in unit of nm. Smaller value means faster rise time in an edge slope, hence sharper edges. The x -axis is pixel location and y -axis is the gray level values. The calculated measure converts to a score in nm unit by multiplying the pixel size, which is 1 nm in this case. We picked this metric since we want the restoration framework to deblur or sharpen the image, especially at the pattern edges.

Since we have 10 runs for each die and each run has our frames, we average these 40 images and use it as reference image when calculating SSIM and PSNR.

2.4 Model Architecture

As computer vision tasks become more challenging over the years, the complexity of neural network architecture has increased drastically to model complex functions. The search space dimension for optimal hyperparameters and training parameters is so high that optimizing in a trial-and-error approach becomes infeasible. Neural architecture search has become a heated research area aiming to provide an automatic way to select optimized architecture.²⁷

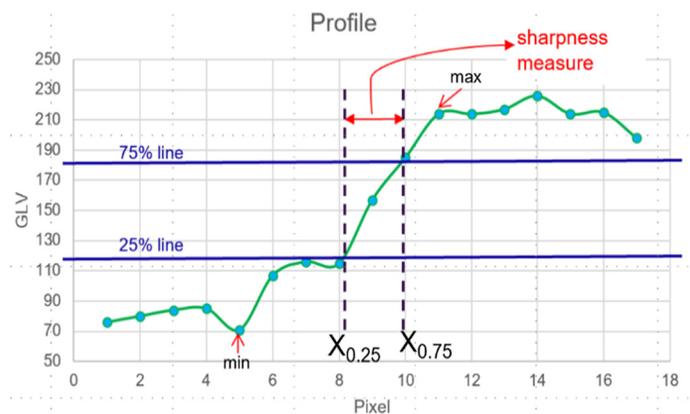


Fig. 6 Sharpness based on 25% and 75% levels between maximum and minimum values on the gray level profile.

Table 4 Neural network architecture under evaluation.

Filter number (n_1, n_2, n_3, n_4, n_5)	(8, 8, 16, 32, 32)
Filter number ratio α	0.5

Table 5 IQ metrics before and after restoration.

	Original 1-frame	Restored 1-frame	Four-frame avg.
PSNR \uparrow	11.88	13.84	12.66
SSIM \uparrow	0.099	0.245	0.260
Sharpness (nm) \downarrow	3.59	2.43	2.51

This approach usually resorts to cross validation performance to guide automatic selection of architecture.

However, due to the lack of ground truth image and self-supervised nature of this methodology, using cross validation to optimize model architecture is infeasible.

Alternatively, since we want to restore the IQ as much as possible while keeping the intrinsic geometrical patterns undistorted, we use the IQ metrics introduced in previous sections to guide architecture selection. The neural network architecture under evaluation is in Table 4. To better balance underfitting/overfitting and maximize the application scenario for different pattern complexity, we picked a network of intermediate complexity. Table 5 shows how all IQ metrics improve significantly after restoration and reach comparable values as the four frames averaged image, suggesting successful restoration. The up-arrow means the higher the value, the better the IQ, and vice versa.

2.5 Model Regularization and Convergence

In this section, we discuss how regularization affects convergence. We use AEI pattern for restoration, we fixed training parameters with input noise standard deviation of 0.01, a fixed learning rate of 0.002, and total 120 iterations. ADAM optimizer was used for all learning processes. It has been discussed previously that we can add prior on image or kernel as a form of regularization so the inverse optimization does not easily converge to trivial or random solutions. We explore how regularization strength on image and kernel TV gradient affects the convergence behaviors. We experiment with different regularization coefficient combinations to better understand how the strength of each regularization term affects convergence as well as learned kernel and restored image.

Figure 7 shows the loss function versus iterations for several selected regularization coefficient combinations to illustrate their effects on learned kernels, and the loss is plotted in log scale for better visualization. We use zero regularization (blue curve) as baseline condition and study the effects of regularization coefficients by varying them individually. The kernel snapshots are at iteration 20, 60, and 120, respectively. When λ_k increases from 1 to 100 (orange to green), kernel becomes sparser and shows pixelation characteristics. This is due to the typical sparsity inducing effect of TV regularization. Nevertheless, over-regularization leads to failure of kernel convergence. When $\lambda_k = 1$, we can observe the kernel becoming less noisy and sparse as iteration increases while still being able to converge to underlying blurring kernel characteristics. Note the kernel size is 21×21 and is artificially enlarged for better visualization. This experiment confirms the divergent nature of this inverse problem and the importance of choosing the appropriate regularization strength.

We further examine if IQ metrics could be used as a proxy for the quality of restored latent images. Figure 8 shows SSIM and PSNR under different image regularization strengths. When $\lambda_x = 20$ (over-regularized), the model converges to a trivial solution containing no pattern, this is because the over regularization strongly prefers image with little features such as edge and corner. This correlates well with low IQ metrics. Comparatively for $\lambda_x = 0.1$ (appropriately

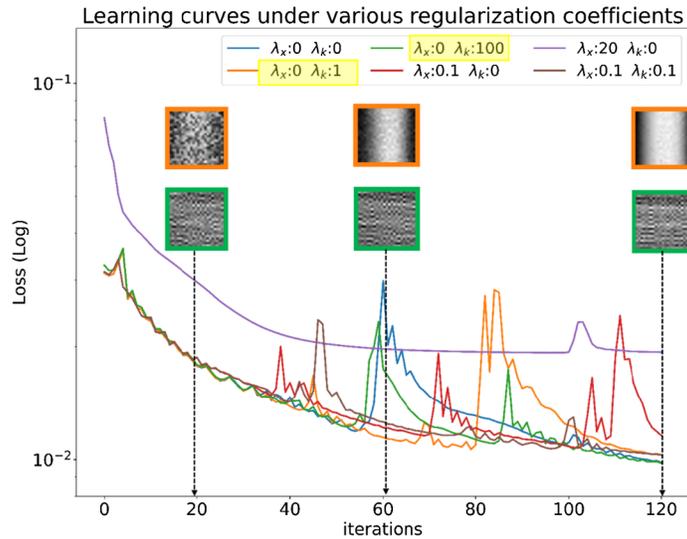


Fig. 7 Loss function versus iterations under different regularization strength combinations.

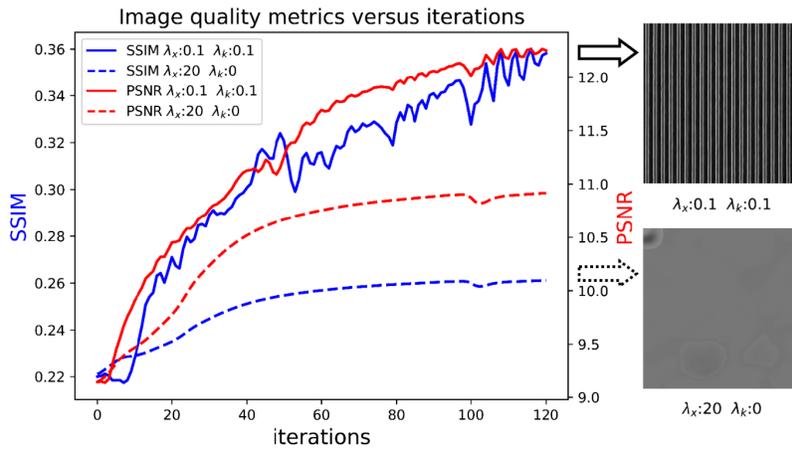


Fig. 8 SSIM and PSNR with iterations under different regularization strength combinations.

regularized), the model successfully restores low noise and sharp latent image with high structural similarity with reference image which correlates well with high IQ metrics. Hence, appropriate regularization strength is crucial to the successful restoration and IQ metrics, such as PSNR and SSIM, could be used to evaluate convergence and determine stopping criteria from an IQ perspective when proper cross-validation technique is not feasible.

3 Results and Analysis

This section analyzes line edge roughness (LER) and critical dimension (CD), which provides us insights as to how image restoration affects metrology results.

It has been established that CD histogram of the image after restoration shows a much smaller standard deviation and suppressed outliers compared to that of low-dosage image, and this was demonstrated in a previous conference paper.²⁸ This is because poor IQ leads to inaccurate CD measurements with higher standard deviation (1.676 nm) and is reduced after restoration (1.057 nm). The restoration also biases the mean CD value by only a small amount (47.08 nm before versus 46.78 nm after). These are metrology results of AEI patterns from Fig. 9.

We conduct LER characterization for left EPE and right EPE using power spectral density (PSD) estimation to further confirm our observations. We place measurement gauges along the direction of the line patterns with a 2 nm interval for both left and right edges, respectively,

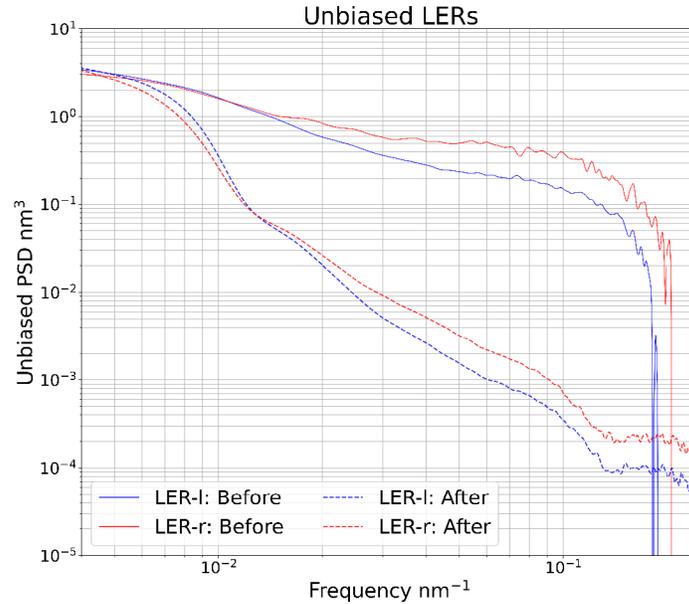


Fig. 9 Unbiased LERs for left and right EPEs before and after image restoration.

to obtain the EPEs, then we average the left and right LERs for all line patterns from all images to get a better estimation. The PSD formula is shown as Eq. (13). f is the spatial frequency, $\Delta d = 2 \text{ nm}$, $x_n = x_0 + (n\Delta d)$ is signals sampled at discrete positions for a total measurement of N samples, we study $\text{Re}(\text{PSD}(f))$ since we care about spectral amplitude. Note that the actual PSD is achieved when N approaches infinity and the expected value can be obtained accurately. In real world cases, the number of real measurement samples is finite hence averaging PSD over many trials is necessary to more accurately estimate the underlying physical process:^{29,30}

$$\text{PSD}(f) = \lim_{N \rightarrow \infty} \frac{(\Delta d)^2}{(2N+1)\Delta d} \left| \sum_{n=-N}^N x_n e^{-i2\pi f n \Delta d} \right|^2. \quad (13)$$

Due to the stochastic nature of LER measurement, we break the variance into three parts as in Eq. (14).³¹⁻³³ $\sigma_{\text{intrinsic}}$ refers to the stochasticity of LER profiles caused by process variation, $\sigma_{\text{metrology}}$ is the measurement uncertainty introduced by metrology algorithm and in our case influenced by IQ, $\sigma_{\text{extrinsic}}$ refers to extrinsic noise contributions from factors such as SEM shot noise, SEM tool stage movement, and beam profile. Unbiasing could be used to remove high frequency extrinsic noise and the results are shown in Fig. 9.³¹ By comparing the unbiased PSD spectrums in Fig. 9, we can see the middle frequency range was drastically attenuated after restoration. This is due to the reduction of metrology noise due to the improvement of IQ. This is in accordance with the observation that CD standard deviation reduces after restoration:

$$\sigma_{\text{psd}} = \sigma_{\text{intrinsic}} + \sigma_{\text{metrology}} + \sigma_{\text{extrinsic}}. \quad (14)$$

Hence, this image restoration method can be used to reduce metrology noise by improving IQ without introducing mean shift and expose intrinsic pattern edge stochasticity. This might be able to provide valuable process information if the underlying process of the patterns is known. The relationship between process and corresponding PSD characteristics is complicated and beyond the scope of this paper and not discussed here.

4 Conclusion

Effective restoration of low-quality SEM images is critical in future high-performance metrology applications as the user pushes for higher throughput and faster turn-around time. This paper introduced a new methodology based on a self-supervised, generative, neural network model. A huge advantage of this approach is that it does not require high-fidelity “ground truth” image

for training, making it especially desirable for low-dosage metrology applications since such ground truth data are usually unavailable. Detailed description of model architecture and regularization was provided. It has been shown by applying the proposed framework IQ can be improved greatly while preserving intrinsic pattern geometry. CD precision, mean CD, and overall distribution confirm the effectiveness in metrology applications. Extension to 2D patterns is also promising, the image is transformed from a state of non-measurable to one that enables reliable metrology results.²⁸ PSD based LER analysis suggests that the restoration method could reduce metrology noise by improving IQ hence exposes intrinsic process-induced line edge profile stochasticity, which is of great value since process stochasticity becomes more prominent as the device feature keeps shrinking in size. We think more use cases of this restoration framework are yet to be discovered.

Code, Data, and Material Availability

The data utilized in this study were obtained from wafers manufactured by our customer, and they do not allow sharing their images freely. The source code that supports the finding of this article is not publicly available because the work is patented and is considered an internal IP for ASML. Nevertheless, to foster collaboration and research, the code could be requested by contacting the author at zijian.du@asml.com

Acknowledgments

This article is based on the conference proceeding paper in 2022 SPIE Advanced Lithography (Ref. 28) and is a detailed extension of it. The authors would like to acknowledge Dr. Rui Yuan for his technical help for conducting the LER analysis and general discussion of research ideas.

References

1. L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.* **79**(6), 745 (1974).
2. S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction methods of multipliers," *Found. Trends Mach. Learn.* **3**(1), 1–122 (2010).
3. W. H. Richardson, "Bayesian-based iterative method of image reconstruction," *J. Opt. Soc. Am.* **62**(1), 55–59 (1972).
4. L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," <http://www.cse.cuhk.edu.hk/leojia/projects/l0deblur/> (2013).
5. A. Chakrabarti, "A neural approach to blind motion deblurring," *Lect. Notes Comput. Sci.* **9907**, 221–235 (2016).
6. S. H. Chan et al., "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.* **20**(11), 3097–3111 (2011).
7. D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR* (2011).
8. R. Wang and D. Tao, "Recent progress in image deblurring," in *SIGGRAPH Asia* (2013).
9. C. J. Schuler et al., "A machine learning approach for non-blind image deconvolution," in *CVPR* (2013).
10. A. Levin et al., "Understanding and evaluating blind deconvolution algorithms," in *CVPR* (2009).
11. J. Kruse, C. Rother, and U. Schmidt, "Learning to push the limits of efficient FFT-based image deconvolution," in *ICCV* (2017).
12. K. Zhang et al., "Plug-and-play image restoration with deep denoiser prior," in *CVPR* (2019).
13. K. Zhang et al., "Learning deep CNN denoiser prior for image restoration," in *CVPR* (2017).
14. S. Xie et al., "Non-blind image deblurring method by the total variation deep network," *IEEE Access* **7**, 37536–37544 (2019).
15. D. Gong et al., "Self-paced kernel estimation for robust blind image deblurring," in *ICCV* (2017).
16. Y. Nan, Y. Quan, and H. Ji, "Variational-EM-based deep learning for noise-blind image deblurring," in *CVPR* (2020).
17. M. Hradis et al., "Convolutional neural networks for direct text deblurring," in *BMVC* (2015).
18. O. Kupyn et al., "DeblurGAN: blind motion deblurring using conditional adversarial networks," in *CVPR* (2018).
19. M. L. Green, "Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm," <https://ww3.math.ucla.edu/camreport/cam02-55.pdf>.
20. C. J. Schuler et al., "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(7), 1439–1451 (2016).

21. X. Xu et al., "Motion blur kernel estimation via deep learning," *IEEE Trans. Image Process.* **27**(1), 194–205 (2018).
22. R. Fergus et al., "Removing camera shake from a single photograph," *ACM Trans. Graphics* **25**(3), 787–794 (2006).
23. A. Levin et al., "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR* (2011).
24. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention ? MICCAI 2015*, Lecture Notes in Computer Science, vol. **9351**, N. Navab et al., Eds., Springer, Cham (2015).
25. A. Shocher, N. Cohen, and M. Irani, "Zero-shot super-resolution using deep internal learning," in *2018 IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, pp. 3118–3126 (2018).
26. D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *ICLR* (2015).
27. T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: a survey," *J. Mach. Learn. Res.* **20**, 1–21 (2019).
28. Z. Du et al., "Low dosage SEM image processing for metrology applications," *Proc. SPIE* **12053**, 1205309 (2022).
29. A. Hiraiwa and A. Nishida, "Spectral analysis of line edge and line-width roughness with long-range correlation," *J. Appl. Phys.* **108**, 034908 (2010).
30. R. Bonam et al., "Comprehensive analysis of line-edge and line-width roughness for EUV lithography," *Proc. SPIE* **10143**, 101431A (2017).
31. L. Pu et al., "Analyze line roughness sources using power spectral density (PSD)," *Proc. SPIE* **10959**, 109592W (2019).
32. C. A. Mack, "Reducing roughness in extreme ultraviolet lithography," *Proc. SPIE* **10450**, 104500P (2017).
33. A. Hiraiwa and A. Nishida, "Spectral analysis of line edge and line-width roughness with long-range correlation," *J. Appl. Phys.* **108**, 034908 (2010).

Zijian Du obtained his PhD in electrical engineering from Arizona State University in 2019 and joined ASML Silicon Valley as senior software engineer. His research interests include applying machine learning and deep learning based techniques to SEM image quality enhancement as well as contour-based edge placement error (EPE) massive-metrology applications for high volume manufacturing (HVM).

Biographies of the other authors are not available.