Denoising sample-limited SEM images without clean data

Lei, Hairong, Teh, Cho, Yu, Liangjiang, Fu, Gino, Pu, Lingling, et al.

Hairong Lei, Cho Teh, Liangjiang Yu, Gino Fu, Lingling Pu, Wei Fang, Hairong Lei, Cho Teh, Liangjiang Yu, Gino Fu, Lingling Pu, Wei Fang, "Denoising sample-limited SEM images without clean data," Proc. SPIE 11611, Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV, 116111A (22 February 2021); doi: 10.1117/12.2584653

Event: SPIE Advanced Lithography, 2021, Online Only

Denoising Sample-limited SEM Images Without Clean Data

Hairong Lei, Cho Teh, Liangjiang Yu, Gino Fu, Lingling Pu, Wei Fang
ASML, 399 W Trimble Rd, San Jose, CA, USA 95131-1028

ABSTRACT

Over the past few years, noise2noise, noise2void, noise2self, and unsupervised deep-learning (DL) denoising techniques have achieved great success, particularly in scenarios where ground truth data is not available or is difficult to obtain. For semiconductor SEM images, ground truth or clean target images with lower noise levels can be obtained by averaging hundreds of frames at the same wafer location, but it is expensive and can result in physical damage to the wafer. This paper's scope is to denoise SEM images without clean target images and with limited image counts.

Inspired by noise2noise, we proposed an additive noise algorithm and DL U-net. We achieved good denoising performance using a limited number of noisy SEM images, without the clean ground truth images. We proposed the "denoise2next" and "denoise2best". We compared generative adversarial network(GAN) generated images and Additive noise images for data augmentation. This paper further quantified the impact of image noise level, pattern diversity, and continuous (aka transfer) learning. The data sets used in the work include both line/space and logic pattern.

Keywords: Denoise, SEM, Additive Noise, Gaussian noise, GAN, Continues learning, Sample-limited, No clean data

1. INTRODUCTION

Recent years have seen deep-learning (DL) based denoisers outperform state-of-the-art conventional denoisers such as the BM3D. They are usually trained to minimize the mean squared error (MSE) between the output image of a DL CNN and a ground truth image or clean target image. Thus, it is important for such DL based denoisers to use high quality clean target or ground truth data for high performance.

The reality is that in many cases it is very expensive or very difficult to obtain clean target images and sometimes it is even impossible (for example medical images or space technology images). Clean semiconductor target images can be obtained by averaging hundreds of frames at the same wafer location, but this can cause permanent physical damage to the wafer for reasons such as resist shrinkage or carbonization. Even when damage is unlikely, there is the acquisition time cost. Figure 1 shows the good image quality enhancement (IQE) results reported in our SPIE2020 paper, where unsupervised deep learning was used to improve the SEM IQE without target images or frame averaging [5].

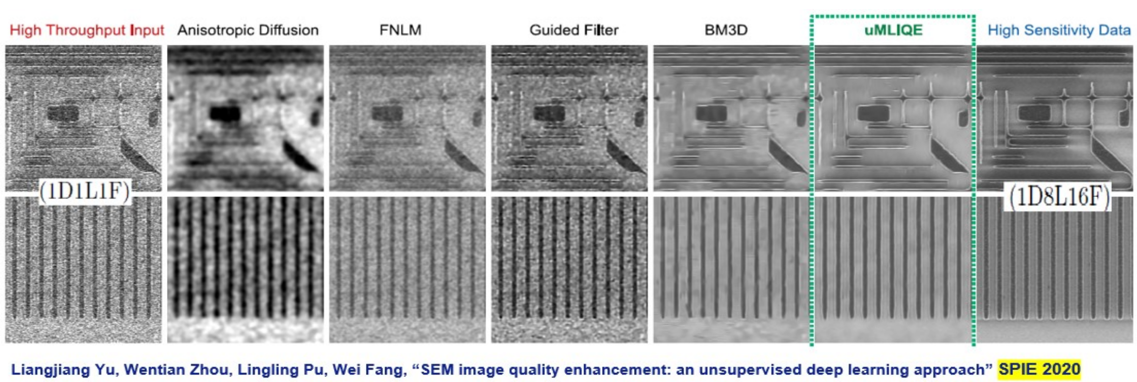

Figure 1: Our 2020 paper using unsupervised DL without frame averaging designed for same condition SEM image capturing.

Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV,
edited by Ofer Adan, John C. Robinson, Proc. of SPIE Vol. 11611, 116111A
© 2021 SPIE · CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2584653

Inspired by the noise2noise [1] concept, we build on this work for the situation where the number of SEM images is limited, with limited frame-average images, and without ground truth images. The use case is different than last year's paper [5]. For denoising performance, the noise2noise is slightly better than noise2void [2] and noise2self [3]. Usually Deep Learning requires many training samples (several hundreds or thousands), but these are not always available, for instance for unique logic patterns. Also training time can be reduced if similar performance is possible with less images. Using our predesigned additive noise algorithm and DL U-net, we are able to achieve good denoising performance using a limited number of noisy SEM images, without the need for clean ground truth images.

DL usually requires large a training set. Generative adversarial network (GAN) generated images have been used in image augmentation for increasing the training set in recent years. We compare the denoise results when using GAN generated SEM images vs. when using our predesigned additive noise algorithm to obtain the training SEM images to train model. We also quantify the impact of image noise level, pattern diversity, and continuous (aka transfer) learning. Both line/space and logic pattern SEM images are used in our DL denoise experiments.

2. ALGORITHM AND EXPERIMENT SETUP

2.1 Noise2noise

Lack of clean data will greatly hinder the performance of an image denoising neural network, because no clean reference is set as the target for the neural network output. This problem was solved by recent research, which demonstrated the feasibility of denoising images using only noisy data. They have mathematically proved the feasibility of this method, using a model called noise2noise [1].

Noise2Noise trains the network using pairs of noisy images of the same wafer area, but independently sampled noise. This work is feasible to learn external image priors and noise statistics from the training data. Noise2Void [2] predicts a pixel from its surrounding pixels by learning blind-spot networks. The work assumes that the noisy image is zero-mean and independent between pixels. However, as mentioned in Noise2Self [3], it significantly degrades the training efficiency and denoising performance at test time. Noise2Self proposes a framework for blind denoising based on self-supervision. They use groups of features whose noise is independent conditional on the true signal to predict one another. This allows them to learn denoising functions from single noisy measurements of each object, with performance close to that of supervised methods.

Theoretically the Noise2Noise and Noie2Self performance is better than that of Noise2Void and unsupervised network. In this paper, we use the DL Noise2Noise, where training targets the best noisy image at hand (averaged frame) without using clean data.

2.2 Predesigned Additive Noise Algorithm

In this paper, we propose a predesigned additive noise algorithm to effectively solve the training image augmentation issue. In our case, we assume the Gaussian noise for the noisy images $y_i = x + n_i$ which are acquired from SEM to train an image-specific network. The n_i is a slightly different noise in (image sequence) 1F, 2F, 3F, …. To reduce the gap on noise statistics, by applying the predesigned noise level n_s, we come up with simulated noisy images $z = y_i + n_s$ with simulated noise n_s, which is statistically close to the noise n_i in y_i. In this way, in the training stage, our U-net learns to remove the simulated noise n_s from z, and thus is able to remove the observed noise n_i from the noisy test image y_i in the testing stage. The model training does not necessary including the test image y_i itself so that the denoise of model could still do a good denoise job for unseen images for the patterns seen in training.

Noise at a range of levels is added to each image in the data set. The purpose is to generate a much larger data set and noise level coverage without losing image structure. The following is algorithm details:

1) We extract noise of image sequence 1F, 2F, 3F…; the noise max level is σ_0 (usually it is from 1F).

2) For each image, we add Gaussian noise distribution from std5, 10, … , k* σ_0. k ~= 2, 3, 4, 5* so that k* σ_0 < 90%.

3) Train the noise2noise CNN model using the above simulated images (denoise to next vs denoise to best).

4) Inference the input noisy images with well-trained model and obtain the denoised image.

2.3 GAN Generated SEM images

A generative adversarial network (GAN) [6] is a class of machine learning frameworks designed by Ian Goodfellow and his colleagues in 2014. The generator generates candidates while the discriminative network evaluates them. The generator tries to fool the discriminator network by producing novel candidates that the discriminator does not realize are synthesized. GAN is a popular data augmentation method. However, GAN has stability problems and can take a long time to train (~4-8 hours in GPU). Furthermore GAN may suffer from a "mode collapse" where it fails to generalize properly or misses entire modes from the input data.

In recent years there has been a lot of interest in domain adaptation using adversarial training[7,8]. The image-to-image translation paper by Isola et al.[8] describes a method that learns to change an image from one domain to another. Basically, given a training set, the GAN learns to generate new data with the same statistics as the training set. For example, a GAN trained on SEM images can generate new SEM images that look at least superficially authentic to human observers. Figure 2 shows the GAN theory and GAN generated SEM image vs. additive noise SEM image:

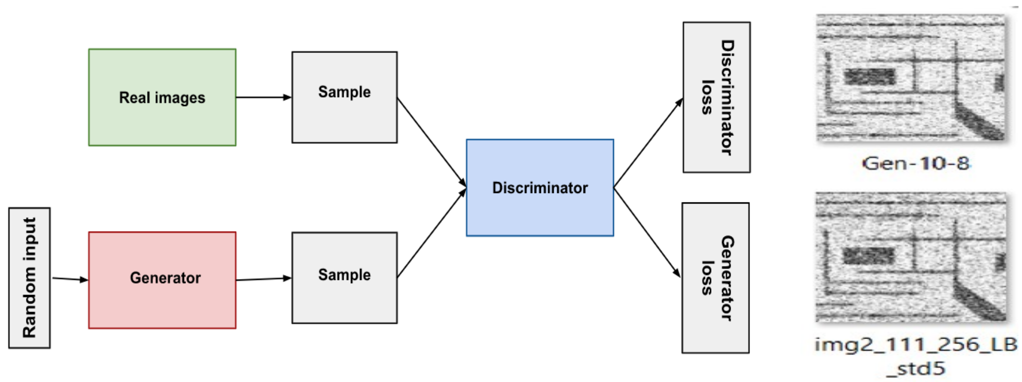

Figure 2. GAN theory and GAN generated SEM image (Gen-10-8) vs. additive noise SEM image (img2_111_256_LB_std5).

2.4 "Denoise2next" vs. "Denoise2best"

In our logic pattern case, we only have four SEM images: img2_111, img2_212, img2_414, and img2_828. The img2_828 is the least noisy image. We call it "denoise2best" (green arrow) when using img2_828 as the training target. The "denoise2next" (red arrow) means that we use img2_212 as the target for img2_111 with additive noise, img2_414 as target for img2_212 with additive noise, …, during training. The detailed difference is shown on the Figure 3.

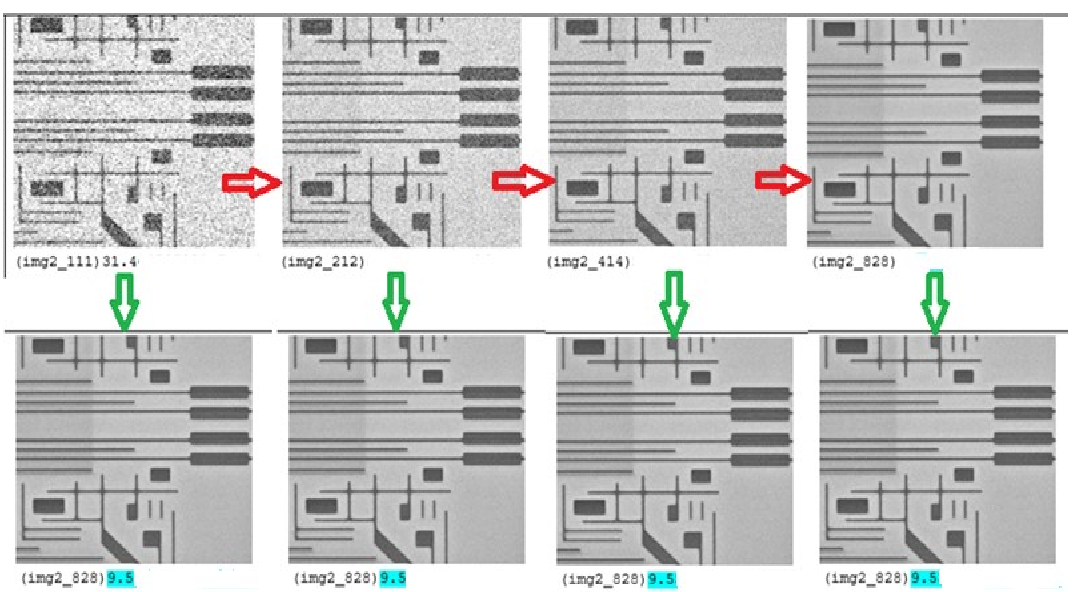

Figure 3. denoise2next (red arrow) vs. denoise2best (green arrow)

2.5 U-net Experiment Setup

Our Noise2Noise U-net uses PyTorch. In training, ADAM optimizer was used with parameter values $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\varepsilon = 10E-8$. We set the learning rate as 0.001, training Minibatch size of 4, input and output channel as 1, input images normalized to [0; 1], with no batch normalization, dropout or other regularization techniques. For logic patterns, we use the img2_111, img2_212, img2_414, and img2_828 as the base with additive noise (img2_828 means 8D2L8F, which means 8 dot 2 line 8 frame averages). For Line/space patterns, we use the image1F, 2F, …, 8F as the base and select some of them with additive noise for our training set. The input image size is either 1024x1024 or 512x512 with cropping size 256x256 for training. The training epochs is set up to 500.

3. EXPERIMENT RESULTS AND ANALYSIS

3.1 Denoise result comparison: GAN or Additive noise, which is better?

This one example shows that our well-trained "predesigned additive noise" model performance is better than that of GAN generated images for denoising SEM image in terms of both PSNR (peak signal to noise ratio) and SSIM (structural similarity index). For inference of img2_111, the GAN model denoise performance is PSNR=35.97 and SSIM =95.4% while the Additive noise model denoise performance is PSNR=37.62 and SSIM =98.9%.

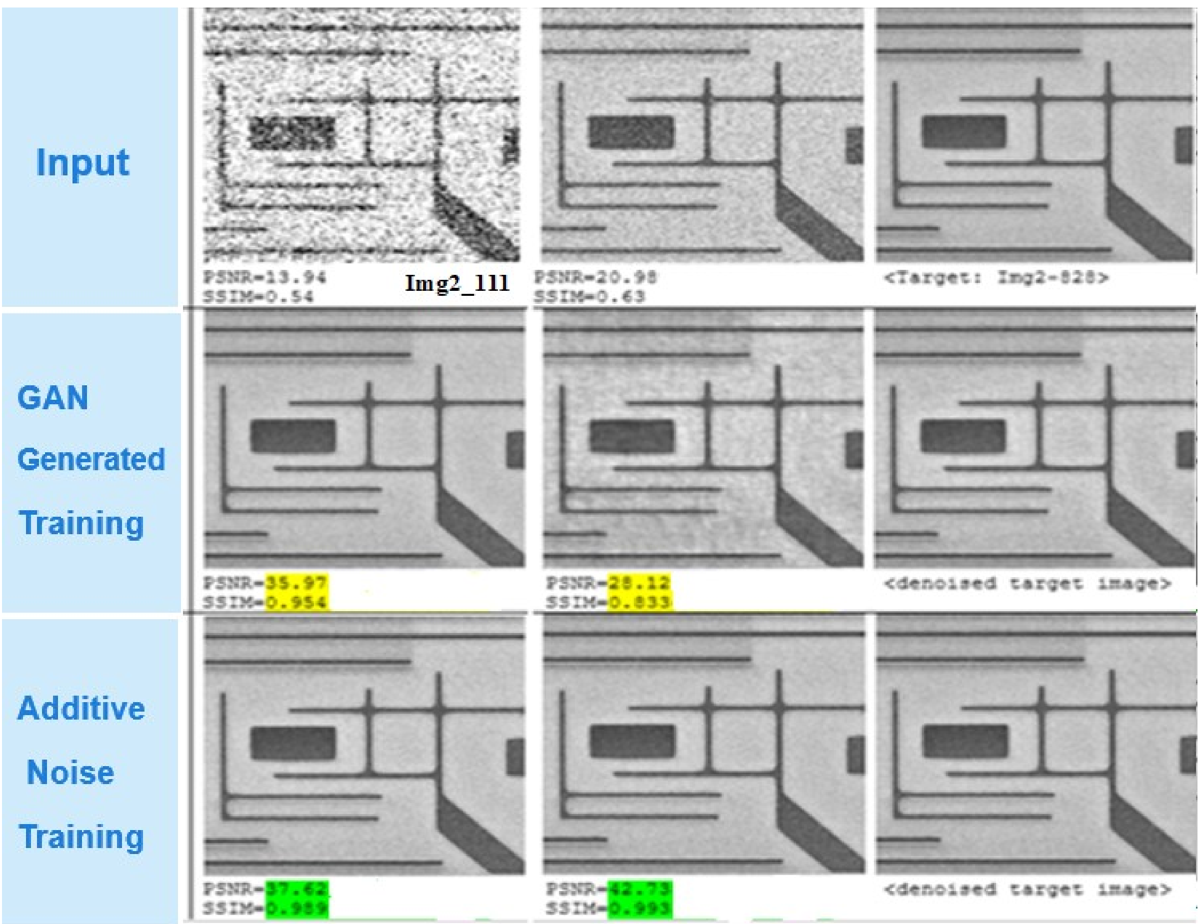

Figure 4. denoise result from GAN generated SEM images vs. Additive noise SEM images

Both models use 32 images in training and train to best target, which is img2_828. There is no ground truth SEM image or clean image. The PSNR and SSIM is using this img2_828, a less noisy image, as the virtual clean image for the image quality metric value calculation.

3.2 PSNR, SSIM, and Noise extraction

For image denoise, the key performance indices (KPIs) including the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [4] as we used in previous section. PSNR and SSIM are popular metrics for image restoration. However PSNR suffers from excessive smoothing, which is very difficult to recognize indistinguishable images. SSIM depends on brightness, contrast and structure, and therefore cannot accurately evaluate image perceptual quality. One may further argue neither PSNR nor SSIM makes much sense when we have no clean image. So we will use extracted noise level σ value to quantify denoise result in this paper. The smaller the σ value, the less noisy the image is. Table 1 shows the extracted Gaussian noise level σ values for image1F to image8F. It is from 15.85 to 5.64. The image32F is our virtual clean image with σ = 2.82. One observation is that the extracted noise level σ value is not linear to the noise number we add as shown in the table to the right in Figure 5. The σ value for image32F_std40 is about 39, which is close to 40 while the σ value for image32F_std70 is not very close to 70. It is only 58, which is a relatively large gap. But this does not prevent us from using it as an image denoising metric for simplicity and effectiveness.

Table 1. SEM Images, additive noised SEM Images, and estimated noise level.

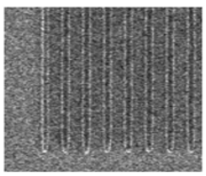

Image1F

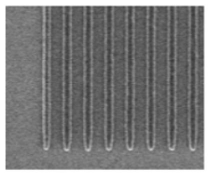

Image8F

Input SEM image	Estimated noise level Sigma σ
Image1F	15.85
Image2F	11.22
Image3F	9.20
Image4F	7.96
Image5F	7.12
Image6F	6.50
Image7F	6.03
Image8F	5.63

Input SEM image	Estimated noise level Sigma σ
image32F	2.81
image32F_std5	5.77
image32F_std6	6.64
image32F_std7	7.53
image32F_std8	8.50
image32F_std9	9.46
image32F_std10	10.41
image32F_std11	11.31
image32F_std12	12.32
image32F_std15	15.25
image32F_std16	16.27
image32F_std20	20.18
image32F_std30	29.76
image32F_std40	39.00
image32F_std50	46.77
image32F_std60	53.34
image32F_std70	58.35

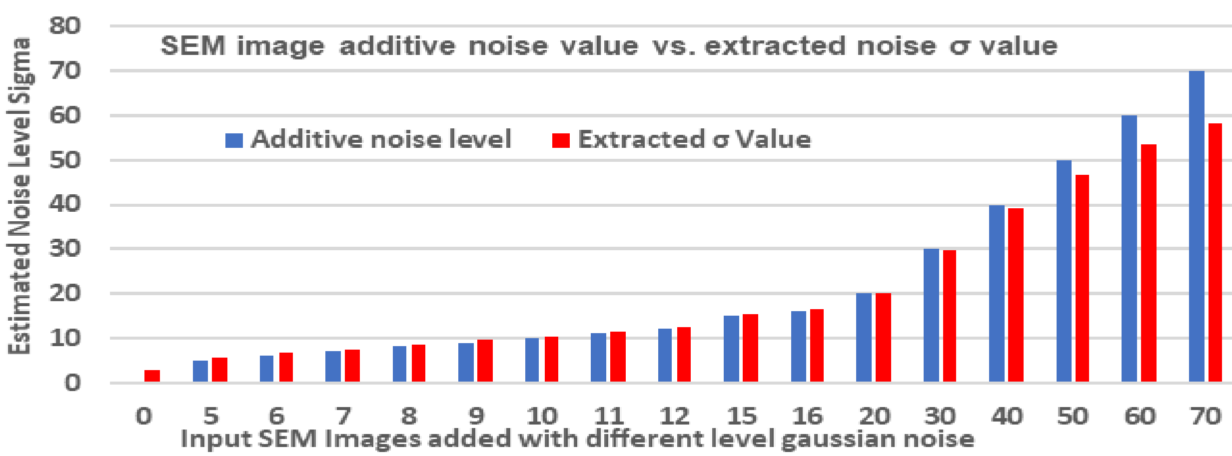

Figure 5. Noise level value deviation (additive noise value vs. estimated/extracted noise level value).

3.3 Does data augmentation help?

Figure 6 shows that line/space pattern denoising result using our additive noise (data augmented) model outperforms that of using only real SEM images to train the model. Both models are trained using "denoise2best", with image32F with σ = 2.82 . When denoising image1F with the Non-additive Model, the extracted noise level σ = 2.57 which is much worse than that of the additive model σ = 1.17. The same trend is observed for the other images in Figure 6.

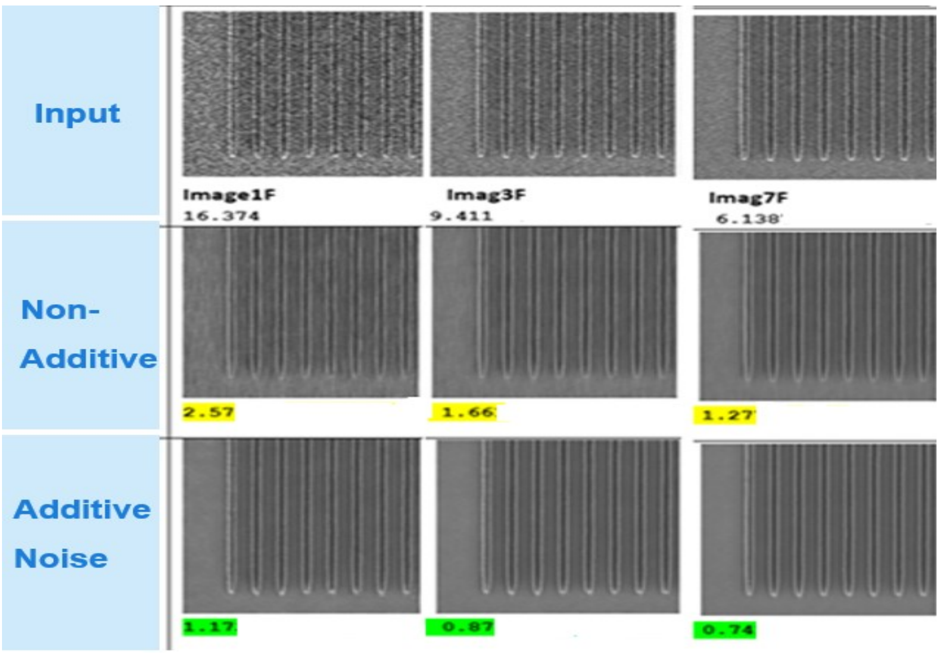

Figure 6. denoise results with model trained by sample-limited SEM images vs. additive noise SEM images

3.4 What noise image is the best target, next or best?

Figure 7 shows that logic pattern denoising result using "denoise2best" outperforms that of using "denoise2next" as mentioned in section 2.4. Both models are trained using our predesigned additive noise SEM images. The training set was augmented to 48 images based on 4 real SEM images .

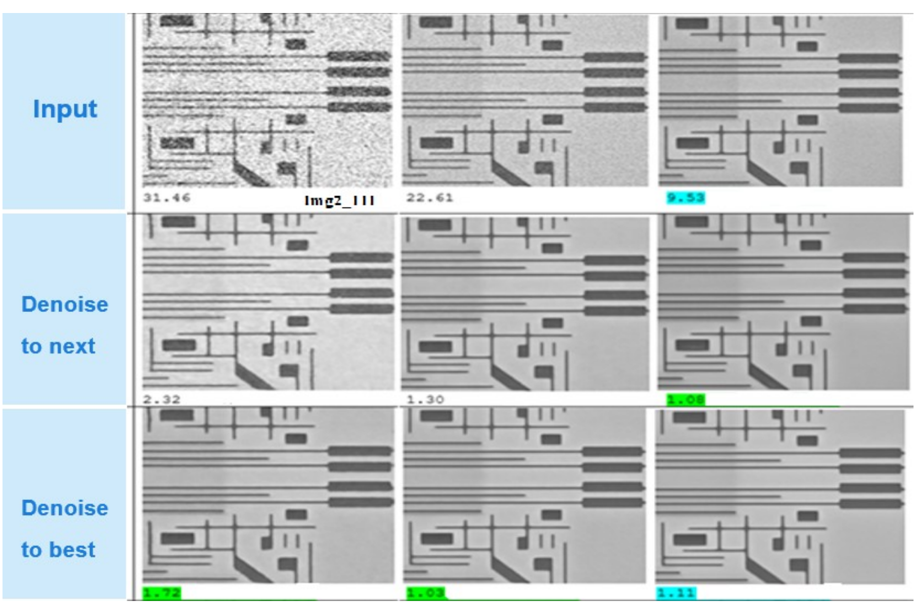

Figure 7. train2next vs. train2best for logic pattern (denoised img2_111 σ =2.32 vs. σ =1.72)

3.5 What is the best range of noise to add?

Noise level coverage plays an important role for denoise model training. For our logic pattern case, the max noise level σ_0 = 31.5; k* σ_0 = 60 ~ 90 when k =2 or 3 with k* σ_0 < 90%. Figure 8 shows the denoise results when training with a noise level range up to σ = 60 ~ 80 is much better than that with a noise level range up to only σ = 40 or σ = 20. From the Img2_111 (top-left corner) denoise result, the denoised image estimated sigma σ is 1.74, 1.72, 3.9, 5.4 when additive noise level coverage with σ = 80, 60, 40, 20 respectively. The σ = (60 ~ 80) is the better than that of σ = (20 ~ 40).

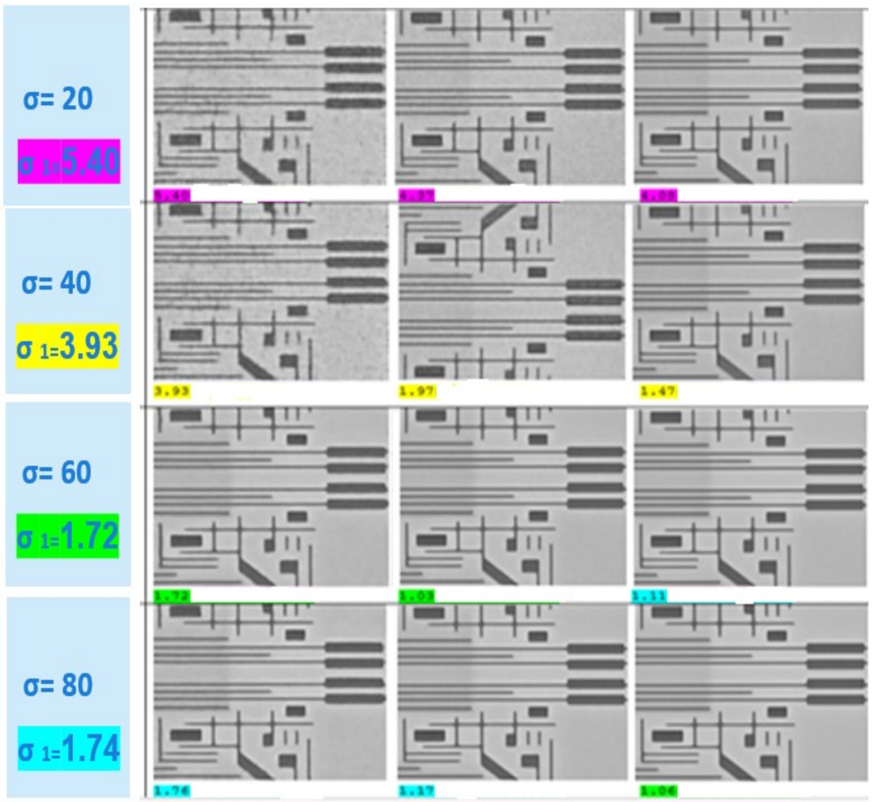

Figure 8. denoise results for model trained with different noise level range in SEM images in training set.

3.6 Does training image count really matter?

The answer is Yes. In Deep Learning, the performance is improved when the data set increased as shown in Figure 9. When we have the same noise level range (σ = 60) in the training set, the denoise performance (**σ1 = 1.72**) of the model that has 48 training images outweighs the performance of the model with 16 images (**σ1 = 2.97**) for img2_111 (top-left).

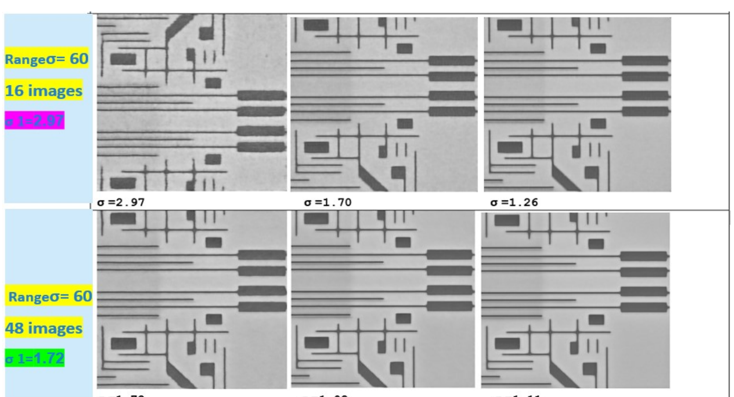

Figure 9. same noise level range but different training image count leads to different denoise performance.

3.7 Can DL trained on one pattern be used for a different pattern?

Issues with pattern coverage are not new for denoising with deep learning. We know that using a denoise model trained on one pattern to denoise a different pattern does not working as shown on Figure 10. Figure 11 shows the denoising performance when the DL denoising model is trained with both patterns. The denoised img2_111 has σ =1.50 and the denoised image1F has σ =1.21 while the denoised img2_414 has σ =1.08 and the denoised image4F has σ =0.87 using the model trained by two patterns.

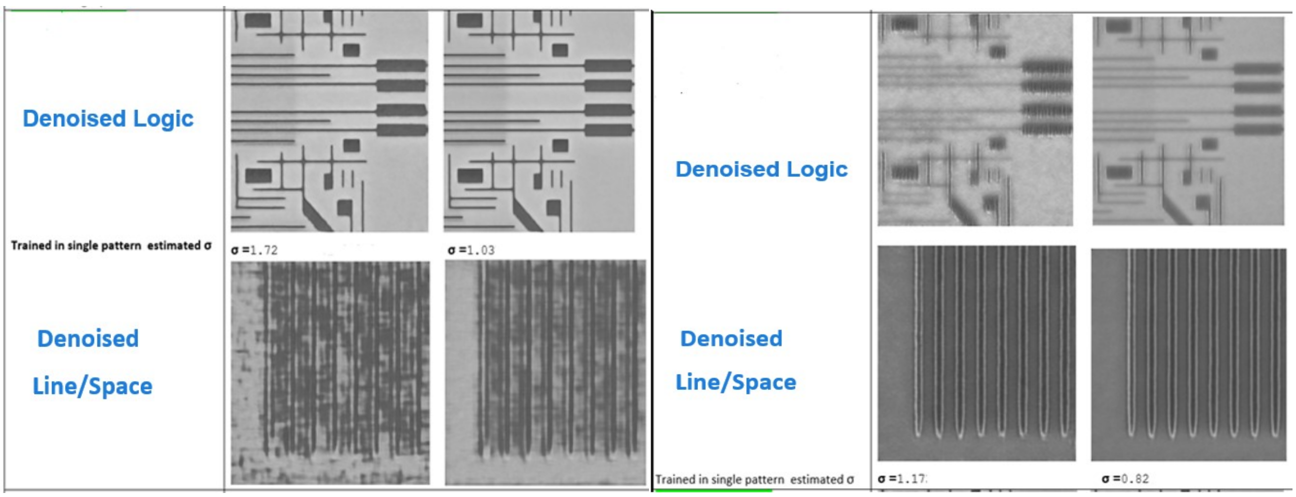

Figure 10. denoise result trained only by one pattern.

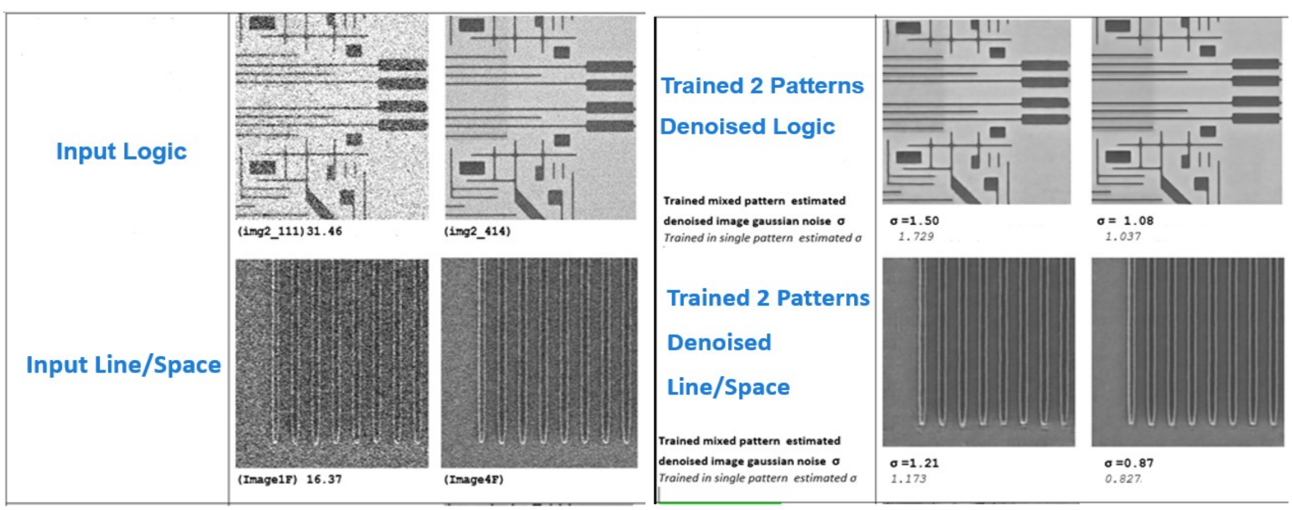

Figure 11. Inputs (left) and denoised logic and line/Space SEM images by model trained in two patterns (right)

3.8 Loss vs. training epochs and training converge time.

The U-net denoise network converges fast. It converges after 40-60 epochs and stabilizes after 80-100 epochs. We train the model to 500 epochs in about 60 mins. The 80-100 epoch is reached in about 10-15 mins. This is true using either the GAN generated SEM images or the additive noise SEM images. An epoch is a full pass over all of training data. A loss is a scalar value that we attempt to minimize during our training of the model. The training set has only 32 images.

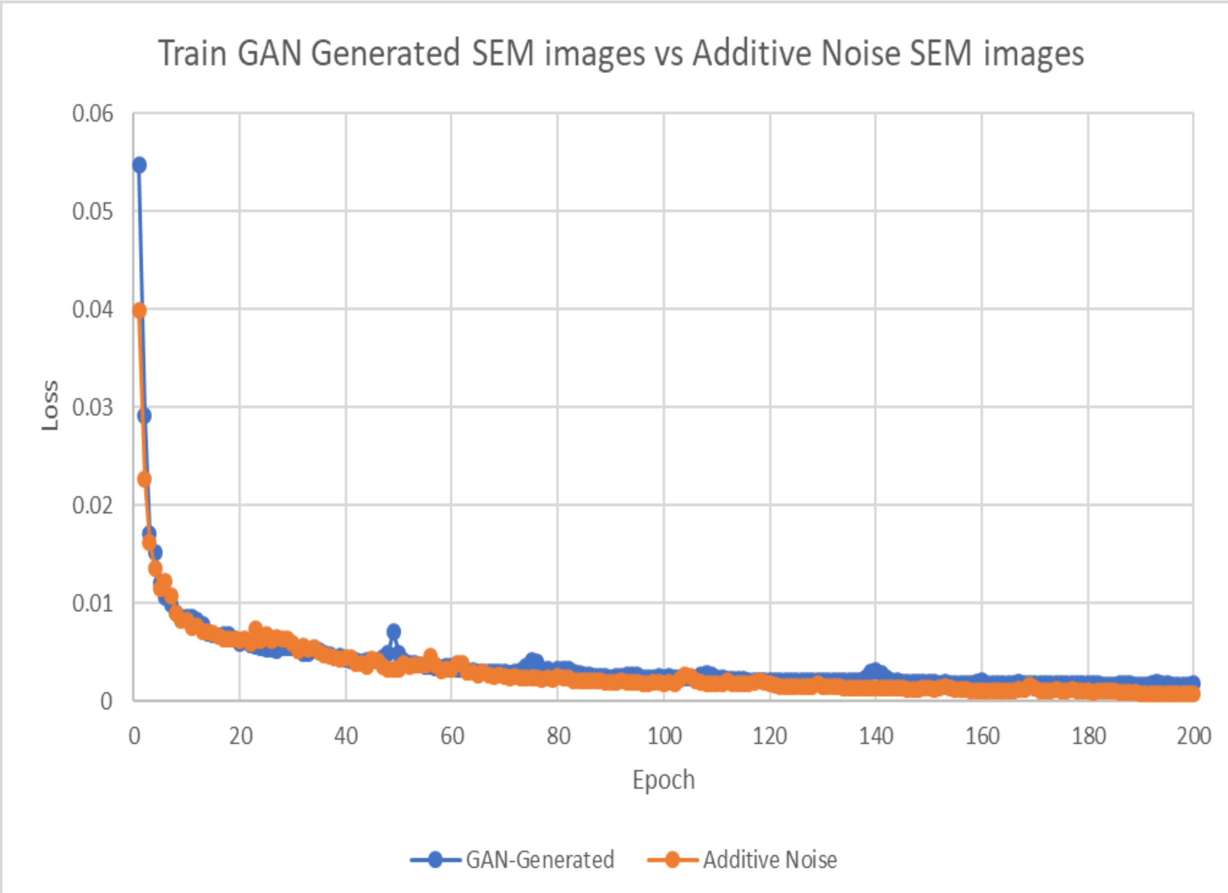

Figure 12. Loss vs. training epochs of U-net for GAN generated and additive noise images

3.9 Continues learning/Transfer learning or Train from Scratch?

Transfer learning is very popular machine learning technique for training classifiers efficiently, where we copy the weights from another trained model before starting the training process for new data. During this process, most of the early layers could be initialized, or frozen while only the last or last several layers would be updated to fine-tune the network. In our application of image de-noising, for a new image pattern image, we apply a similar scheme, that is to continuously learn our model for the new pattern. This could be done to save a lot of training resources. As shown in Figure 14, we have tested this method of training on a logic pattern. The first row are inference results of model trained only on logic pattern, while the second row are inference results from a model continuously trained leveraging a line pattern model. Notice that on epoch 1, the continuous trained model has superior edge sharpness. As the training progresses, from epoch 1 to epoch 3, 5, and 7, we still observe the advantage of continuous training in terms of noise level and edge sharpness. Such effects would be more obvious when huge amounts of patterns could be used for the model learning. So continuous training could be potentially very helpful for inline machine learning image de-noising throughput.

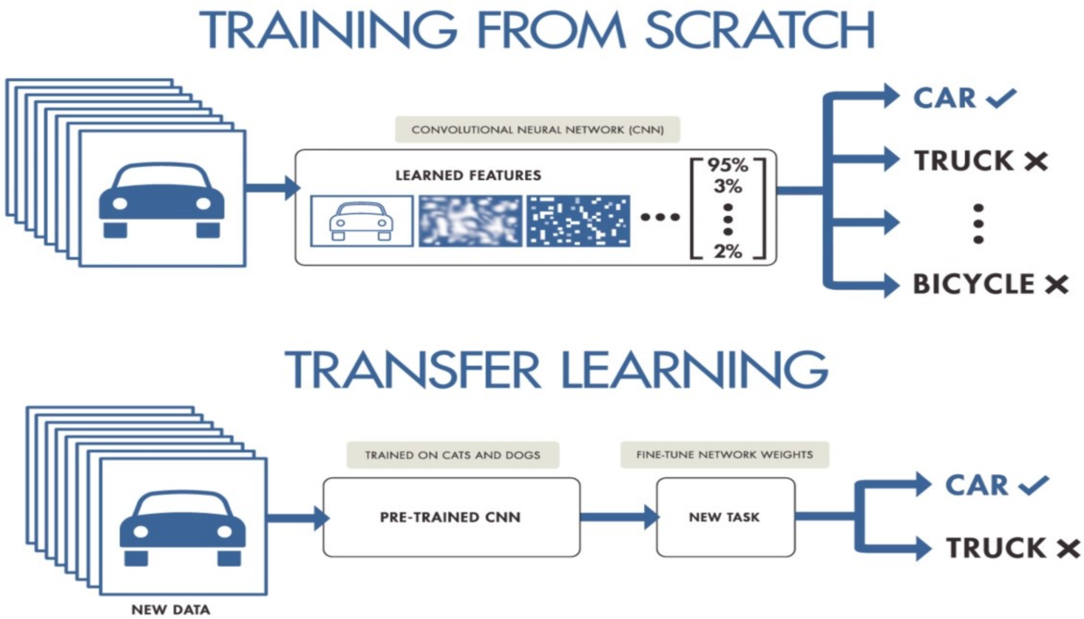

Figure 13 Training from Scratch vs. Transfer Learning (classification)/Continuous Learning (denoising) [10]

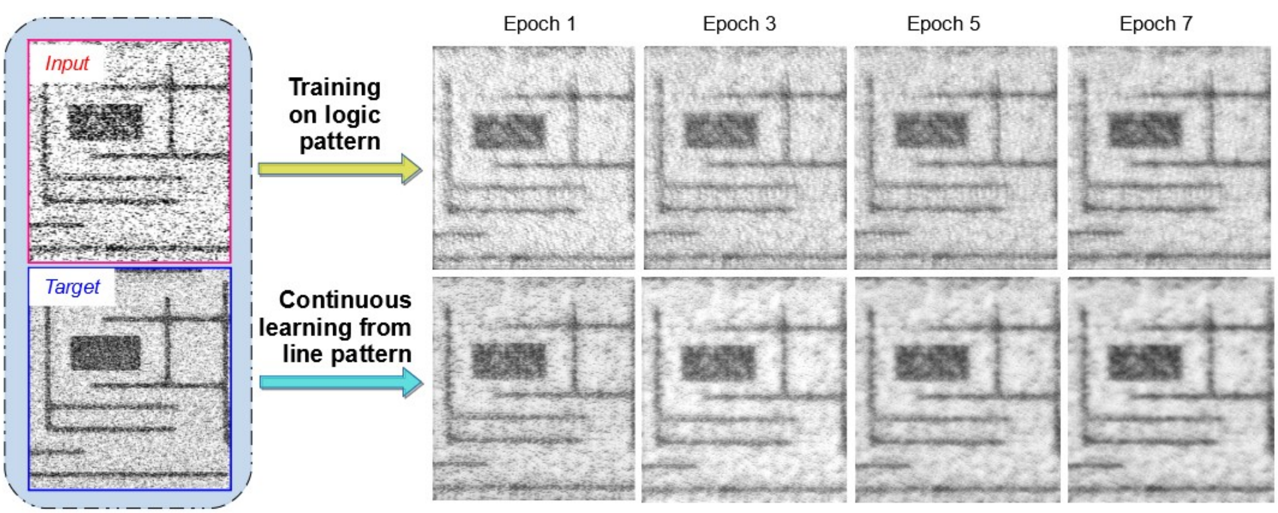

Figure 14 IQE qualitative comparison on logic pattern. The first row is results from ML model trained solely on logic pattern, while the second row is the inference from a model initialized based on the training process using line patterns, and continuously trained on logic pattern afterwards. The advantage is more obvious when there are many more patterns.

4. SUMMARY

We explained our noise2noise denoising U-net framework to handle sample limited SEM image (4- 8 SEM images) denoising. We proposed a pre-designed additive noise algorithm based on image extracted noise level when clean data is not available. We compared our additive noise training denoise results with GAN generated SEM image training denoise results. We studied the impact of training noise level coverage and pattern coverage in the training set. We provided suggestions on how to select the noise level coverage for the training set. We showed our U-net framework has good converge and only needs 80-100 epochs training (training time 10-15 mins). Continuous training could be potentially very helpful for inline machine learning image de-noising throughput (2x to 8x throughput improvement).

ACKNOWLEDGEMENT

We really appreciate the assistance of Oliver Patterson, Gary Zhang, Ronald Goossens, Paul Hinnen, Stefan Hunsche, Jan Mulkens, Bob Socha, Jun Lang, Marc Kea, and Koen Thuijs from ASML who reviewed this paper's content and provided helpful suggestions. We also want to thank Kangsheng Qiu and Yufei Duan's or ASML for providing us with the experimental SEM data.

REFERENCES

[1] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T., "Noise2Noise: Learning image restoration without clean data," in [Proceedings of the 35th International Conference on Machine Learning], Proceedings of Machine Learning Research 80, 2965-2974, PMLR (10-15 Jul 2018).

[2] Krull, A., Buchholz, T., and Jug, F., "Noise2void - learning denoising from single noisy images," in [IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019], 2129-2137 (2019).

[3] J. Batson and L. Royer. "Noise2Self: Blind denoising by self-supervision". In ICML, volume 97, pages 524–533. PMLR, 2019

[4] Zhou Wang, Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing 13, 600-612 (April 2004).

[5] L. Yu, et al., "SEM image quality enhancement: an unsupervised deep learning approach, Proceedings of SPIE Advanced Lithography Conference, Feb 2020.

[6] Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). Generative Adversarial Networks (PDF). Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

[7] Salimans, Tim; Goodfellow, Ian; Zaremba, Wojciech; Cheung, Vicki; Radford, Alec; Chen, Xi (2016). "Improved Techniques for Training GANs". arXiv:1606.03498

[8] Isola, Phillip; Zhu, Jun-Yan; Zhou, Tinghui; Efros, Alexei (2017). "Image-to-Image Translation with Conditional Adversarial Nets". Computer Vision and Pattern Recognition.

[9] Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. "U-net: Convolutional networks for biomedical image segmentation". MICCAI, 9351:234–241, 2015.

[10] https://medium.com/@1154_75881/transfer-learning-628e83df5c8a