Enhancing Semiconductor X-RAY Images: A Framework Combining Denoising and Super-Resolution Modules With a Novel Dataset

Jae Hoon Shim*[‡], Min Woo Kim*[‡], Tae Gyu Lim[†], Byung Seok Min[†], Sang Hwa Lee*[§], Nam Ik Cho*

* Department of ECE, INMC, Seoul National University, Seoul, Republic of Korea

[†] Central R&D Center, XAVIS Co., Ltd., Seongnam, Kyeonggi-do, Republic of Korea

[‡] indicates equal contribution

§ corresponding author

Abstract—Semiconductors are essential components of modern technology, and inspecting their defects has become increasingly important. One popular approach is to inspect them using X-ray images, but the release of semiconductor images to the public is restricted due to security concerns. This limited dataset poses a challenge, particularly when developing learning-based inspection methods. Another problem with X-ray images is the noise and low resolution, which lowers the inspection accuracy of machine and human inspectors. To overcome this challenge, we introduce a new public semiconductor dataset and propose a new X-ray image enhancement framework to help inspectors increase inspection accuracy. Precisely, we have collected a new dataset consisting of three types of X-ray images obtained from actual industrial settings. Additionally, our enhancement method can improve all types of images without the need to train a separate model for each type. This is achieved by analyzing the noise within the dataset and building on previous studies, and using objective functions defined in the frequency domain. Various experiments and ablation studies have shown that our method can significantly improve X-ray semiconductor images, both qualitatively and quantitatively. Our proposed semiconductor dataset is publicly available at this URL: https://drive.google.com/drive/folders/1_ YSuxR_YGNah1mnPGNd-pgrRULNZP5gn?usp=sharing.

Index Terms—Image Enhancement, Semiconductor, X-ray Imaging, Image Denoising, Image Super Resolution

I. INTRODUCTION

Semiconductors are important components used in various products and manufacturing machines today. They are also critical parts of machines where safety is a top priority, such as unmanned vehicles. Therefore, accurately detecting defects in semiconductors is becoming increasingly important, and improving the speed and accuracy of defect inspection is a critical objective in the semiconductor industry.

One popular semiconductor inspection method is to identify defects through X-ray imaging [1]–[4]. In general, the precision of semiconductor inspection is largely dependent on the quality of X-ray images, where two key factors that affect the image quality are the *energy level* and *exposure time* [5]. Although higher energy levels and longer exposure times can enhance the image quality, they can also have negative effects.

Precisely, it is required to have a delicate balance between image clarity and potential damage to the semiconductor circuits and other structures. While higher energy levels result in clearer images, excessive energy can harm the semiconductors. Therefore, it is crucial to limit the energy level to avoid damage. In addition, the exposure time should be long enough to produce clean images with minimal noise from the equipment. However, a long exposure time can also reduce semiconductor yields. Hence, to obtain high-quality images without harming the semiconductors, short exposure time with limited X-ray energy is necessary. But this can negatively impact the accuracy of both human and machine inspectors.

Therefore, in this paper, we present a new semiconductor image dataset and a deep-learning approach to improve the quality of X-ray images captured with various energy levels and exposure times. Since most semiconductor images are intended for industrial purposes, their public release is strictly prohibited. This restriction has resulted in a scarcity of datasets containing semiconductor X-ray images, which in turn has hindered the advancement of image processing technologies related to semiconductors. As demonstrated in Figure 1, our dataset includes three types of semiconductor images captured under various equipment and conditions, and we expect that the dataset can help the academic community overcome the research limitation.

Furthermore, based on prior research [6] and our observation of commonalities in the noise distribution across them, we suggest an integrated framework to enhance input images, instead of training separate networks for each type. Our framework consists of two modules: a denoising module and a super-resolution module. Initially, to reduce noise, we employ downsampling in the spatial domain using an encoder. Subsequently, due to the importance of preserving essential details like circuit and bump boundaries on semiconductor wafers, we adopt the architecture proposed in Soh et al. [7]. In addition to the spatial domain loss, we incorporate frequencydomain functions, Focal Frequency Loss [8] and Waveletdomain High-Frequency Loss [9], which are shown to keep high-frequency components. As a result, our method shows significant improvements in both qualitative and quantitative ways.

Our contributions can be summarized as follows:

• We have proposed a new semiconductor X-ray image



Fig. 1. Overview of our proposed dataset. The dataset consists of three types of data, each having a different resolution $(S_1 \in \mathbb{R}^{1024 \times 1024}, S_2 \in \mathbb{R}^{1936 \times 1536}, S_3 \in \mathbb{R}^{1900 \times 1900})$. Low-Quality images have significant noise due to short-duration X-ray exposure, while High-Quality ones have less noise as they are projected over a comparatively longer duration.

dataset that comprises three types of images captured under different equipment and conditions. The aim of this dataset is to facilitate research in the academic community, particularly learning-based X-ray image processing research.

• We have demonstrated that there are common statistical characteristics among the types of data. Based on the analysis, we designed an integrated framework capable of improving all three types at once.

• Through experiments and ablation studies, we have validated that our approach makes a significant improvement in both qualitative and quantitative results.

II. METHOD

A. Novel Semiconductor Dataset

We propose a novel semiconductor X-ray image dataset. As depicted in Figure 1, to provide diversities, our dataset includes three distinct types of images, depending on the filming equipment and environment. These are denoted as S1, S2, and S3. Each type comprises pairs of Low Quality (LQ) and High Quality (HQ) images. LQ images are captured by subjecting semiconductor wafers to a brief X-ray exposure (5 seconds), resulting in images with significant noise. Meanwhile, HQ images are produced by a longer X-ray exposure (90 seconds), yielding images with considerably less noise. There are 2135 images in S1, 2000 in S2 and 1040 in S3 dataset, and we have partitioned each data type into train and test datasets. For S1 and S3, the proportion between training and test datasets is 80:20, while it is 90:10 for S2.



Fig. 2. To investigate the noise characteristics added to Low Quality (LQ) images compared to High Quality (HQ) ones, we plot histograms of the pixel values obtained by subtracting LQ from HQ, using 100 training pairs from each data type. Across all three data types, the shapes of histograms are similar to Gaussian.

B. Theoretical Background for the Framework

1) Noise Analysis in Dataset: In order to examine the nature of noise in LQ images relative to HQ ones, we compute the differences between corresponding pairs and analyze the pixel histogram for each type of data. As shown in Figure 2, we find that the additional noise present in LQ images across all types (S1, S2, and S3) can be approximated by a Gaussian distribution and exhibits a mathematically similar form. There have been many studies [6], [10], [11] for handling Gaussian noise of various standard deviations using a single model. This research provides the theoretical foundation for our integrated model.

2) Noise2Noise: Noise2Noise [6] suggested that it is possible to train the network to produce clean data solely with the use of noisy image pairs. They demonstrated that if the noise expectation is zero, the model can learn to approximate the clean ground truth despite each gradient aligning with the noisy data. The optimization task can be formulated as:

$$argmin_{\theta} \sum_{i} L(f_{\theta}(\hat{X}_{i}), \hat{Y}_{i}) where \ \hat{Y}_{i} = Y_{i} + n_{i}.$$
(1)

In this context, θ represents the trainable parameters of the model, and the pair (\hat{X}_i, \hat{Y}_i) denotes a set of noisy training images. Note that Y_i and n_i indicate the clean data and noise, respectively. Under the assumption that the noise has a zero mean, such as Gaussian noise, the network's parameters are optimized to generate an unobserved clean target as:

$$E[\hat{Y}_i|\hat{X}_i] = Y_i \tag{2}$$

especially when the L2 loss is utilized. Furthermore, employing the L1 loss eliminates outliers, which similarly guides the network's training process, resulting in less blur in the output images.

Our goal is to convert data with noise caused by brief Xray exposure times into clean data. However, obtaining clean



Fig. 3. Overall structure of the proposed framework. The LQ image, targeted for enhancement, initially goes through the Denoising Module. This process results in the denoised feature where unnecessary information like noise is eliminated. However, it retains sufficient information for the Super-Resolution Module in the feature ($z \gg c$ in Equation 3). Then, the feature undergoes the Super-Resolution Module to produce an Enhanced-Quality image. Utilizing Enhanced Quality and HQ images, the final objective function (Equation 4) is calculated to train our framework.

semiconductor X-ray images theoretically needs infinite X-ray exposure time, which is physically unfeasible. Thus, inspired by Noise2Noise research, we intend to project input data onto the manifold of clean data using pairs of noisy images, involving more noisy images (Low Quality, LQ) and less noisy ones (High Quality, HQ).

C. Framework

The primary objective of our framework is to enhance LQ images to achieve clarity comparable to HQ ones while eliminating noise. Hence, our framework consists of two key modules: the denoising module and the super-resolution module, as illustrated in Figure 3.

Traditional techniques for noise reduction involve low-pass filters and bicubic downsampling. However, with the advancement of deep learning, it has been proven that an end-to-end trained network outperforms these traditional methods in denoising task [12]–[16]. Especially, the downsampling process within such networks is critical for removing high-frequency bands that contain noise components [17]. Consequently, we incorporate an encoder network with downsampling in the denoising module to effectively filter out high-frequency noise from LQ images.

To ensure semiconductor images perform optimally in postprocessing stages, such as anomaly detection, it is important not only to eliminate noises but also to improve image clarity. However, in the spatially compressed feature obtained through downsampling via an encoder, there is a lack of high-frequency information, which is related to the image detail. Hence, when enlarging the spatial size of the embedding feature to match the original image, we utilize a network designed for the Super-Resolution (SR) task. Within our framework, we adopt the fractal residual learning (FRL) architecture proposed in NatSR [7]. Although NatSR uses a discriminator to determine if the input patch lies on the natural manifold, we only utilize the SR network for processing semiconductor X-ray images because the discriminator's role is not proper in our task.

The above series of processes can be written as the following equation:

$$\hat{y} = D(f) = D(E(x)), \tag{3}$$

where $x \in \mathbb{R}^{h \times w \times c}$ and $\hat{y} \in \mathbb{R}^{h \times w \times c}$ indicate LQ and Enhanced Quality (EQ) image. Additionally, $f \in \mathbb{R}^{h/4 \times w/4 \times z}$ represents the denoised feature, which is the most spatially compressed element within our framework. E and D denote the denoising and the SR module, respectively. Note that in contrast to Singh et al. [18], we assign a larger value to z in fthan to c in x. This assignment allows us to incorporate more information into f, which is essential for the SR module to effectively restore image details.

D. Objective Function

To train the network in the spatial domain, we utilize a distortion-oriented loss function. We refrain from adopting losses associated with the discriminator, as our goal is to produce the unobserved clean data manifold. Furthermore, since this field does not involve human visual perception and aesthetic appreciation, the use of any perceptual loss is deemed less relevant.

To improve the clarity of semiconductor lines and bump boundaries in images, we incorporate two losses defined in the frequency domain. First, we utilize Focal Frequency Loss (FFL) [8], suggested to narrow the frequency gap between EQ and HQ images. This method converts each image to the frequency domain via the Discrete Fourier Transform (DFT) and then calculates the Euclidean distance and associated weight at each spectral position. Nevertheless, these weights are proportional to the scale of the frequency band, assigning the low-frequency band with larger weights. To emphasize the high-frequency Loss (WHFL) [9]. This approach ensures that weights increase with frequency, effectively highlighting higher frequencies.

As a result, our final objective function for training the network can be written as:

$$L(y, \hat{y}) = |y - \hat{y}|_{1}^{1} + \lambda_{1} \cdot L_{FFL}(y, \hat{y}) + \lambda_{2} \cdot L_{WHFL}(y, \hat{y}),$$
(4)

where y and \hat{y} denote HQ and EQ images, respectively. Also, λ_1, λ_2 are hyper-parameters balancing among each loss term.



Fig. 4. Comparison of results among original Low Quality (LQ), High Quality (HQ) images and super-resolved images according to downsampling scale $(DS_{n\times})$. $DS_{1\times}$ means that the input is not downsampled at all. $DS_{2\times}$ and $DS_{4\times}$ indicate that the input goes through $2\times$ and $4\times$ downsampling, respectively.

TABLE I

EXPERIMENTAL RESULTS ON DIFFERENT SETTINGS FOR THREE DATA TYPES: S1, S2, AND S3. THE EXPERIMENTS COVER VARIOUS SETTINGS, INCLUDING "BICUBIC" (APPLYING BICUBIC DOWNSAMPLING TO THE INPUT), "DM" (DENOISING MODULE), "FFL", AND "WHFL". PERFORMANCE RANKINGS ARE INDICATED WITH RED DENOTING THE HIGHEST, AND BLUE INDICATING THE SECOND-HIGHEST PERFORMANCE FOR EACH DATA TYPE. NUMBERS IN PARENTHESES REPRESENT THE PERFORMANCE INCREASE COMPARED TO LQ IMAGES.

Setting	DM	FFL	WHFL	Data type	PSNR	Average PSNR	SSIM	Average SSIM
				S1	34.91(+3.22)		0.9264(+0.2218)	
bicubic	X	X	X	S2	33.76(+5.63)	34.39(+6.39)	0.8603(+0.1895)	0.8996(+0.3373)
				S3	33.94(+13.62)		0.8824(+0.7158)	
				S1	37.20(+5.52)		0.9296(+0.2250)	
bicubic + FFL	X	\checkmark	X	S2	33.79(+5.67)	35.60(+7.60)	0.8600(+0.1893)	0.9010(+0.3387)
				S3	34.04(+13.71)		0.8819(+0.7154)	
				S1	35.29(+3.60)		0.9310(+0.2263)	
DM + FFL	\checkmark	\checkmark	X	S2	33.97(+5.84)	34.80(+6.80)	0.8659(+0.1952)	0.9036(+0.3413)
				S3	34.62(+14.29)		0.8838(+0.7172)	
DM + FFL + WHFL	~	\checkmark	\checkmark	S1	35.34(+3.65)		0.9324(+0.2278)	
				S2	33.91(+5.78)	34.83(+6.83)	0.8662(+0.1955)	0.9048(+0.3424)
				S3	34.68(+14.36)		0.8851(+0.7186)	

III. EXPERIMENT

A. Experimental Settings

We conduct experiments to compare four different settings across three data types: S1, S2, and S3, using PSNR and SSIM as evaluation metrics. In the "bicubic" setting, we apply bicubic downsampling to the input image and perform superresolution to obtain the resulting image, employing the L1 loss as the loss function. In the "bicubic + FFL" setting, the network architecture remains the same, but we adopt FFL in addition to the L1 loss. In the "DM + FFL" setting, we replace bicubic downsampling with the denoising module. Lastly, in the "DM + FFL + WHFL" setting, we combine both FFL and WHFL with the L1 loss.

B. Experimental Results

Table I presents the PSNR and SSIM results for the four settings. Our proposed framework, "DM + FFL + WHFL," exhibits the best performance across all data types in terms of SSIM. It achieves the second-highest PSNR results in S1 and S2 data types and the highest PSNR in S3. While "bicubic + FFL" demonstrates the highest average PSNR value, this is primarily attributed to the high PSNR value in S1. However, a detailed analysis reveals that each component we introduced (denoising module, FFL, and WHFL) effectively enhances both PSNR and SSIM. Note that we set the SR module as the learning-based method [7] because the module is an essential component for upsampling the downsized intermediate results back to the original image size, excluding the component from this analysis.

Firstly, comparing "bicubic" and "bicubic + FFL," we observe improvements across all aspects, especially a significant increase in PSNR for S1. This suggests that FFL can compensate for the L1 loss for improving the quality of result images.

Secondly, "DM + FFL" enhances most metrics across nearly all data types when compared to "bicubic + FFL." Notably, "DM + FFL" employs a deeper feature, rather than a downsampled image, as an intermediate feature. This approach more effectively preserves the information necessary for restoring image details. Although increasing the channel dimension of the feature may adversely affect the performance in PSNR of S1, it improves overall performance including SSIM in S1. Hence, replacing "bicubic" with "DM" enhances the results.

As mentioned, the "DM + FFL + WHFL" setting, which incorporates WHFL into "DM + FFL", demonstrates superior performance. While maintaining PSNR values in S2, WHFL enhances all other metrics across every data type in comparison



Fig. 5. Comparison among the original LQ, HQ images, and results from existing methods and ours for every data type (S1, S2, S3). Each row corresponds to each data type, and each column indicates an image category. Columns (d) and (e) show the results from EDCNN [19] and NAFNet [20], respectively. Column (f) displays the results of our full framework.

to "DM + FFL". As a result, given that SSIM is based on structural information, we can conclude that both FFL and WHFL significantly contribute to enhancing the quality of X-ray images for semiconductors.

C. Effects of Downsampling

We adopt downsampling as a preprocessing step for input images to eliminate noise and other unnecessary elements before proceeding to the SR module. Therefore, to investigate the effect of downsampling, we conduct an ablation study. In this section, we employ bicubic downsampling to the Low-Quality (LQ) input and subsequently super-resolve it using our deep network.

Figure 4 illustrates the results of applying $1 \times$ (indicating no downsampling), $2 \times$, and $4 \times$ downsampling to the S1-type image. The image downsampled at $4 \times$ shows better line preservation and more efficient noise removal compared to the $1 \times$ and $2 \times$ downsampled images, leading to clearer patterns. Despite the noise in the HQ image, downsampling contributes to obtaining an image cleaner than the HQ image.

Based on the above observation, we conclude that the downsampling process effectively eliminates unnecessary information, such as noise and stain-like artifacts, while preserving essential features like lines and curves. Thus, we design our framework by combining a denoising module for the downsampling process and the SR model to improve image quality, akin to an autoencoder.

D. Comparison to Existing Denoising Methods

We conduct a comparative analysis of our method against established denoising techniques. NAFNet [20] demonstrates exceptional performance in real-world image denoising, while EDCNN [19] works effectively in the context of CT image denoising, particularly in the medical image domain. We have re-trained both NAFNet and EDCNN on our dataset.

As illustrated in the qualitative results of Figure 5, the images denoised with EDCNN exhibit both blurriness and noise. Meanwhile, NAFNet achieves clearer results than EDCNN, but the residual noise artifacts still remain. Notably, more pronounced noise patterns are seen in the results of S2 and S3 data types. In contrast, our method produces clean denoised images with significantly reduced noise artifacts compared to other denoising methods. With our framework, images can be structurally reinforced, enhancing defect inspection capabilities.

IV. CONCLUSION

We have introduced a new dataset and a deep learning framework to improve the quality of semiconductor X-ray images. The dataset includes X-ray images taken at various energy levels and exposure times. In our deep learning framework, the input image is first passed through a denoising network that downsamples the image while preserving important information by expanding the feature's dimensionality. The feature is then processed by a super-resolution network. To train the network, we combined pixel and frequency-domain functions to form a loss function. Our proposed architecture and these functions work together to enhance the quality of the output image, outperforming one of the leading deep networks in the field of real image denoising. We believe that our dataset and framework will contribute to the advancement of X-ray image processing for semiconductors and other areas.

ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program (ATC+ program, 20014131, 25nm X-ray Inspection System for Semiconductor Backend Process) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea), in part by BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2023, and partially by Samsung Electronics Co., Ltd.

REFERENCES

- Q. Zhang, M. Zhang, C. Gamanayake, *et al.*, "Deep learning based defect detection for solder joints on industrial x-ray circuit board images," in 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), IEEE, vol. 1, 2020, pp. 74–79.
- [2] J. O'Leary, K. Sawlani, and A. Mesbah, "Deep learning for classification of the chemical composition of particle defects on semiconductor wafers," *IEEE Transactions* on Semiconductor Manufacturing, vol. 33, no. 1, pp. 72– 85, 2020.
- [3] K. Kim, H. Kim, J. Chun, M. Kang, M. Hong, and B. Min, "Real-time anomaly detection in packaged food xray images using supervised learning," *Comput. Mater. Contin*, vol. 67, no. 2, pp. 2547–2568, 2021.
- [4] J. Ding, C. Ye, H. Wang, J. Huyan, M. Yang, and W. Li, "Foreign bodies detector based on detr for highresolution x-ray images of textiles," *IEEE Transactions* on *Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023.
- [5] I. O. Romero, Y. Fang, and C. Li, "Correlation between x-ray tube current exposure time and x-ray photon number in gate," *Journal of X-ray science and technology*, vol. 30, no. 4, pp. 667–675, 2022.
- [6] J. Lehtinen, J. Munkberg, J. Hasselgren, *et al.*, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.
- [7] J. W. Soh, G. Y. Park, J. Jo, and N. I. Cho, "Natural and realistic single image super-resolution with explicit natural manifold discrimination," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8122–8131.
- [8] L. Jiang, B. Dai, W. Wu, and C. C. Loy, "Focal frequency loss for image reconstruction and synthesis," in *Proceedings of the IEEE/CVF International Conference* on Computer Vision, 2021, pp. 13919–13929.

- [9] M. W. Kim and N. I. Cho, "Whfl: Wavelet-domain high frequency loss for sketch-to-image translation," in *Proceedings of the IEEE/CVF Winter Conference on applications of computer vision*, 2023, pp. 744–754.
- [10] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [11] M. El Helou, R. Zhou, and S. Süsstrunk, "Stochastic frequency masking to improve super-resolution and denoising networks," in *Computer Vision–ECCV 2020:* 16th European Conference, Glasgow, UK, August 23– 28, 2020, Proceedings, Part XVI 16, Springer, 2020, pp. 749–766.
- [12] Z. Tu, H. Talebi, H. Zhang, et al., "Maxim: Multiaxis mlp for image processing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5769–5780.
- [13] M. Chang, Q. Li, H. Feng, and Z. Xu, "Spatialadaptive network for single image denoising," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, Springer, 2020, pp. 171–187.
- [14] S. Lefkimmiatis, "Universal denoising networks: A novel cnn architecture for image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3204–3213.
- [15] Z. Hong, X. Fan, T. Jiang, and J. Feng, "End-to-end unpaired image denoising with conditional adversarial networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34(04), 2020, pp. 4140–4149.
- [16] C. Ren, X. He, C. Wang, and Z. Zhao, "Adaptive consistency prior based deep network for image denoising," in *Proceedings of the IEEE/CVF conference on computer* vision and pattern recognition, 2021, pp. 8596–8606.
- [17] A. H. Ribeiro and T. B. Schön, "How convolutional neural networks deal with aliasing," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 2755– 2759.
- [18] P. Singh, M. Diwakar, R. Gupta, *et al.*, "A method noisebased convolutional neural network technique for ct image denoising," *Electronics*, vol. 11, no. 21, p. 3535, 2022.
- [19] T. Liang, Y. Jin, Y. Li, and T. Wang, "Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising," in 2020 15th IEEE International Conference on Signal Processing (ICSP), IEEE, Dec. 2020. DOI: 10.1109/icsp48669. 2020.9320928. [Online]. Available: http://dx.doi.org/10. 1109/ICSP48669.2020.9320928.
- [20] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European Conference on Computer Vision*, Springer, 2022, pp. 17–33.