Generative High-Magnification Image Synthesis for EUV Mask Inspection

1st Hyejin S. Kim
Artificial Intelligence Robot Research Division
Electronics and Telecommunications Research Institute
Deajeon, Korea
marisan@etri.re.kr

2nd Seongjin Choi
APS Dvision
FST
Dongtan, Korea
sjchoi@fstc.co.k

***Abstract*—This paper proposes a mask inspection image enhancement method based on the pix2pix architecture to synthesize high-magnification images from low-magnification images of EUV semiconductor pattern masks. To the best of our knowledge, this is the first work leveraging generative models to transform low-magnification EUV mask pattern images into their high-magnification equivalents. The proposed method improves defect detection accuracy while reducing the inspection costs associated with high-resolution imaging.**

I. Introduction

Extreme Ultraviolet (EUV) lithography masks are critical for patterning next-generation semiconductor devices, and ensuring these masks are defect-free is essential for chip yield and performance. Detecting tiny pattern defects on an EUV mask typically requires high-magnification imaging (e.g., high-resolution scanning electron microscope views) to reveal fine details that low-magnification scans might miss. However, capturing high-magnification images across an entire mask is impractical due to the limited throughput and high cost of specialized actinic or electron-beam inspection tools. EUV mask inspection equipment remains extremely expensive, and data acquisition is limited, as EUV masks cost approximately 3–10 times more than deep ultraviolet (DUV) masks. This creates a challenge: how to leverage the speed of low-magnification imaging while still achieving the defect visibility of high-magnification views.

Recent advances in generative AI offer a potential solution. Generative models have revolutionized image synthesis and translation tasks in computer vision. In particular, image-to-image translation frameworks can learn to transform an input image into a new image domain, altering resolution or style while preserving content. Generative Adversarial Networks (GANs) marked a milestone in this field, enabling realistic image generation and domain translation [1].

The pix2pix model [2] first demonstrated that a conditional GAN could effectively map paired images from one domain to another (e.g., sketches to photos), achieving impressive results. Since then, numerous GAN-based methods have emerged for tasks such as super-resolution (enhancing image detail) and style transfer [3]–[5]. GAN-based approaches, including super-resolution GANs such as SRGAN [6] and ESRGAN [7], can generate high-frequency details that standard interpolation methods cannot. Diffusion models have recently emerged as state-of-the-art generative models, offering improved image fidelity and training stability compared to GANs [8].

For instance, denoising diffusion models and latent diffusion models (LDMs) iteratively refine noise to generate photorealistic images and have been scaled to very high resolutions. Cutting-edge diffusion-based systems like Stable Diffusion [9] and DALL-E 2 [10] achieve remarkable results but require massive training datasets on the order of billions of images and extensive computational resources [11]. However, in specialized domains such as semiconductor mask inspection, obtaining such large datasets is impractical and costly. Transformers have also been adapted for generative modeling; for example, Diffusion Transformers (DiTs) integrated into diffusion models effectively capture global context for high-resolution image generation [8]. Nonetheless, these sophisticated architectures tend to be data-intensive and complex to train, posing significant challenges given the scarcity of EUV mask training data.

In this work, we address the challenges described above by leveraging a generative image-translation approach to convert low-magnification EUV mask images into synthetically generated high-magnification images. By doing so, we aim to combine the coverage and speed of low-magnification scans with the detailed visibility of high-magnification scans, thereby improving defect detection performance. We adopt a pix2pix-based conditional GAN architecture for this task, rather than transformer- or diffusion-based models, precisely because our EUV mask image dataset is limited. The pix2pix framework—with its encoder–decoder U-Net generator and patch-based discriminator—is particularly well-suited for learning from relatively small paired datasets [5].

Furthermore, GANs enable the model to synthesize fine textural details resembling actual high-magnification patterns through an adversarial loss that encourages outputs indistinguishable from true high-magnification images. By training on a modest number of paired low- and high-magnification examples, our method learns to super-resolve and enhance mask images specifically tailored to the characteristics of EUV patterns. We hypothesize that utilizing these synthetically enhanced images in defect analysis can boost the detection of

Corresponding author: marisan@etri.re.kr

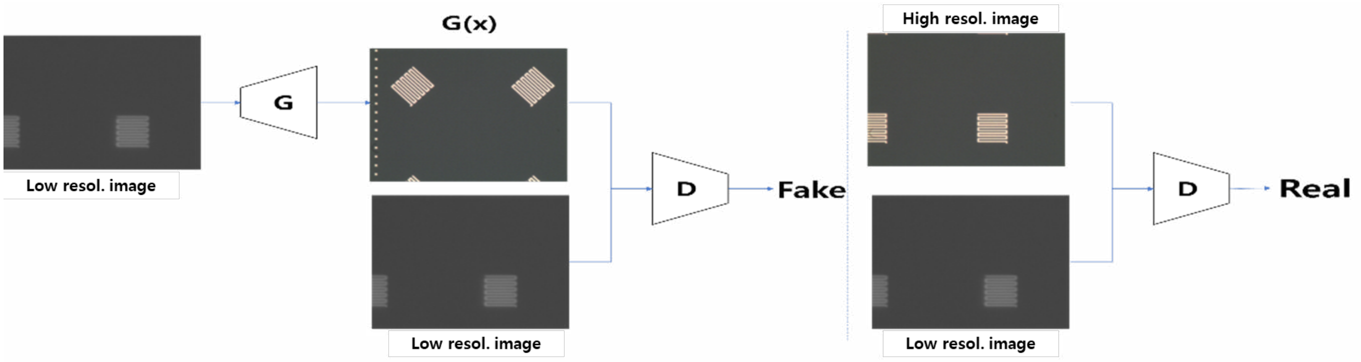


Fig. 1: The overall architecture of the proposed algorithm for generating a high-magnification image from a low-magnification image is presented.

subtle defects that might otherwise be missed at low resolution.

In summary, our contribution is a generative AI-based approach for EUV mask inspection that predicts high-magnification images from low-magnification inputs, facilitating improved defect identification without requiring exhaustive high-magnification scanning.

II. Related Works

A. Image-to-Image Translation and Super-Resolution

Image translation using deep generative models has become a vibrant area of research. Pix2pix introduced the seminal idea of employing conditional GANs for supervised image-to-image translation, learning a mapping from an input domain to an output domain using paired training images [2]. Building on this, pix2pixHD [12] and other variants extended conditional GAN (cGAN) [13] translation capabilities to high-resolution outputs by employing coarse-to-fine generators and multi-scale discriminators, enabling photorealistic results for large images.

In the field of image super-resolution, SRGAN [6] applied adversarial training to upsample images, recovering realistic textures unattainable by pixel-wise loss functions alone. An improved version of SRGAN, ESRGAN [7], introduced architectural and loss-function enhancements (e.g., Residual-in-Residual Dense Blocks and a relativistic discriminator) to produce even more natural image details. These advancements demonstrated that GAN-based super-resolution can effectively generate high-frequency details (e.g., sharp edges, surface textures), making upscaled images perceptually convincing.

However, GANs often suffer from issues such as training instability and mode collapse, especially when generating very high-resolution images [14]. This limitation has prompted exploration into alternative generative approaches, notably diffusion models. Diffusion probabilistic models [15] have recently set new standards in image-generation fidelity. These models generate images by gradually denoising random noise, thereby avoiding mode collapse by more comprehensively covering the data distribution. Latent Diffusion Models (LDMs) further improve efficiency by performing diffusion in a compressed latent space, significantly reducing memory requirements and processing time while preserving image quality. Diffusion models have also been applied to image-to-image tasks. For instance, by conditioning on input images or embeddings, diffusion models can be guided for tasks such as style transfer or super-resolution. Saharia et al. [16] demonstrated diffusion-based super-resolution (SR3), and more recent studies have combined diffusion with transformers for paired image-translation tasks. The DenoSR method employs a pretrained diffusion model for zero-shot super-resolution, effectively enhancing noisy images by progressively refining high-frequency details through inverse diffusion [17].

Although these approaches achieve impressive results, their applicability to our problem is limited by data constraints. Training large diffusion or transformer-based models typically requires very large and diverse image datasets; for example, Stable Diffusion was trained on billions of web images. In contrast, our task addresses a niche data domain—dark-field images of EUV mask patterns—where only a relatively small corpus of paired low- and high-magnification samples is available. This motivates our choice of the pix2pix GAN framework, which has been shown to perform effectively even with limited data and can be efficiently trained on a single GPU within a reasonable timeframe, while still producing sharp, realistic results through its learned adversarial loss.

B. Generative Models in Semiconductor Inspection

Recent research has explored generative models within semiconductor inspection contexts. Zhang and Ma [18] developed a pixel-to-pixel GAN model to accelerate lithography mask simulations. By incorporating deformable convolutions and an LSTM module into a pix2pix framework, they achieved significantly faster prediction of detailed mask diffraction patterns compared to traditional physics-based simulations. Their work demonstrates that conditional GAN architectures can effectively capture complex pattern transformations by mapping mask layouts to aerial images with high accuracy.

In defect inspection, generative networks have been employed both for data augmentation and direct defect detection. For instance, Mohammed and Clarke [19] utilized a conditional GAN for fabric defect inspection, generating synthetic defective images to augment small training datasets. Their model took as input a defect mask and a clean fabric image, outputting realistic composites with defects, thus increasing defect variability for model training. This approach significantly improved defect recognition accuracy by overcoming data scarcity. Generative augmentation methods have similarly been proposed in other industrial inspection contexts to create examples of rarely encountered defects.

Another related line of research involves using neural networks to reconstruct or characterize mask defects from imaging data. Zheng et al. [20] introduced a GAN-driven framework to infer the 3D profile parameters of buried phase defects in EUV mask blanks. By feeding the network multi-angle aerial images of a defect and training it to reconstruct the defect's profile, they achieved high-accuracy characterization that outperformed certain physics-based methods. Their use of GAN models underscores the potential of generative approaches for addressing inverse problems in lithography inspection.

Our work is inspired by these successes and is, to our knowledge, the first to apply an image-translation GAN for EUV mask magnification enhancement. Specifically, we focus on translating low-magnification inspection images into a higher-magnification domain. This can be viewed as a form of super-resolution or domain enhancement tailored specifically to mask inspection. While classic super-resolution research often deals with natural images or straightforward downsampling scenarios, the mapping between low- and high-magnification EUV mask images involves not only higher pixel density but also distinct noise characteristics and imaging artifacts.

By training on true low/high-magnification image pairs from real masks, our model implicitly learns to synthesize realistic fine pattern details (including noise textures) representative of genuine high-magnification images, thereby producing outputs that closely mimic actual high-magnification inspections. We build upon the pix2pix cGAN paradigm, incorporating several domain-specific adjustments to ensure the generated high-magnification images are structurally consistent with their low-magnification inputs. This careful approach prevents the introduction of false defects while visually resembling authentic high-magnification views. The overall architecture of our proposed method is illustrated in Fig. 1.

In summary, the related work on generative image translation and prior applications of GANs in inspection provides a strong foundation for our research. Our approach merges these ideas to fill a critical gap in EUV mask inspection—enabling efficient defect analysis by synthetically zooming in on low-resolution scans—and contributes a novel application of conditional GANs in semiconductor inspection. The following sections describe our model architecture and training procedure in detail, and evaluate the extent to which the generated high-magnification images improve defect detection performance in practice.

III. The Proposed Method

Fig. 1 illustrates the overall architecture of the proposed pix2pix-based generative adversarial network (GAN) employed to synthesize high-resolution EUV mask images from low-resolution inputs. The generator (G) first transforms the low-resolution input image into a corresponding synthetic high-resolution image. Subsequently, the discriminator (D) evaluates pairs comprising the generated high-resolution image and the input low-resolution image against pairs of real high- and low-resolution images. Through adversarial training, the discriminator learns to distinguish between real and synthesized image pairs, thereby guiding the generator to produce increasingly realistic high-resolution images. This conditional adversarial framework enables accurate synthesis of fine details, significantly enhancing defect detection performance in EUV mask inspection.

Since low-resolution and high-resolution images were acquired using different objective lenses (with $10\times$ and $50\times$ magnification, respectively), directly inputting them into the GAN was impractical due to their dimensional mismatch. To overcome this challenge, we introduced a preprocessing stage employing a template matching technique. Specifically, we aligned the image pairs by identifying corresponding regions between the low-magnification ($10\times$) and high-magnification ($50\times$) images, and extracting these matching areas for effective pairing. This preprocessing step ensured consistent input dimensions for the GAN, significantly enhancing the reliability and quality of the synthesized high-resolution images.

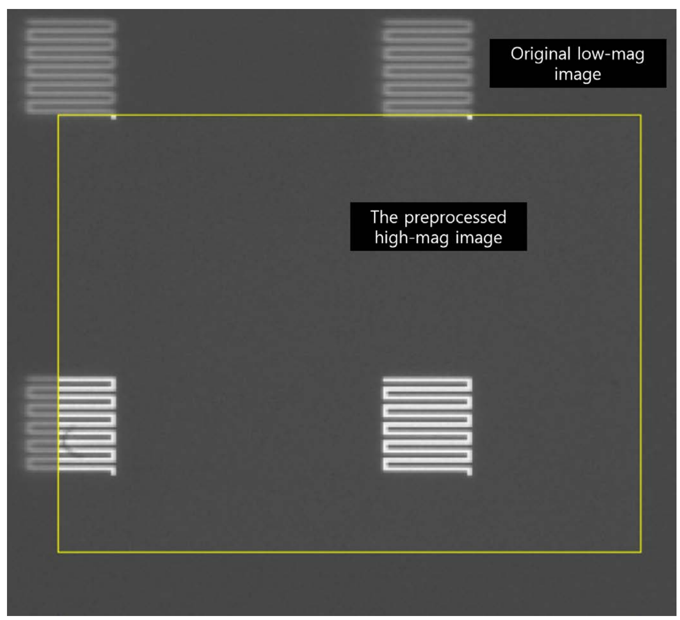


Fig. 2: To verify the preprocessing step for the low-magnification and high-magnification images, we overlaid these images.

Fig. 2 illustrates an example of successful alignment between high-magnification ($50\times$) and low-magnification ($10\times$) images achieved through template-matching preprocessing. First, the original high-magnification image was downsampled to match the dimensions of the low-magnification image. Subsequently, template matching was employed to accurately identify the corresponding region of the high-magnification image within the low-magnification frame. The yellow bounding box indicates the matched region, clearly demonstrating precise spatial alignment. This preprocessing step facilitates consistent image pairing, enabling the GAN model to effectively learn the mapping from low- to high-resolution domains.

IV. Experimental Results

In this study, we employed extreme ultraviolet (EUV) mask images provided by FST Company to construct our dataset and evaluate the matching accuracy between low- and high-magnification image pairs. The low-magnification dataset consisted of 42 images, each approximately 1 GB in size, acquired under a broad field of view. For detailed inspection, we selected 500 regions of interest—comprising both defect and defect-free areas—and captured the corresponding high-magnification images using a review-level optical system.

To standardize the input for our matching algorithm, all 500 high-magnification images were downscaled to 200 × 200 pixels. We then extracted the corresponding regions from the low-magnification images by cropping to the same 200 × 200 dimensions. Ensuring precise alignment between the two magnification levels was critical; therefore, we combined conventional template matching with edge-based feature extraction. Specifically, we first identified candidate regions using normalized cross-correlation, then refined the matches by aligning prominent edge contours extracted with a Canny filter. This two-step approach reduced misalignment and improved pairwise correspondence.

Finally, the assembled dataset of 500 matched image pairs was partitioned into training and testing subsets using a 7:3 ratio, yielding 350 pairs for model training and 150 pairs for performance evaluation. This split enabled a robust assessment of our matching pipeline, ensuring that both defective and non-defective instances were represented in each subset.

To illustrate both the successful and failure modes of our image generation pipeline, we present qualitative examples in Fig.3 and Fig.4. Fig.3 demonstrates that the proposed method can accurately generate complex and realistic EUV mask patterns. Two representative error cases are shown in Fig.4. In each subfigure, the ground-truth high-magnification image and mask are overlaid, with the generated output displayed in semi-transparent red.

Fig. 4(a) shows a case where the model correctly reconstructs one of two adjacent mask patterns (left) but fails to generate the second pattern (right) in the correct location. The blue overlay highlights the missing contours on the right-hand side, indicating that the network sometimes under-represents less prominent pattern elements when multiple features coexist within a single crop.

Fig. 4(b) illustrates an error caused by rotational symmetry in the EUV mask: the model reproduces a pattern that is a 180° rotated version of the ground-truth pattern. Although the overall shape is preserved, this rotation error results in the misalignment of defect positions that are dependent on pattern orientation.

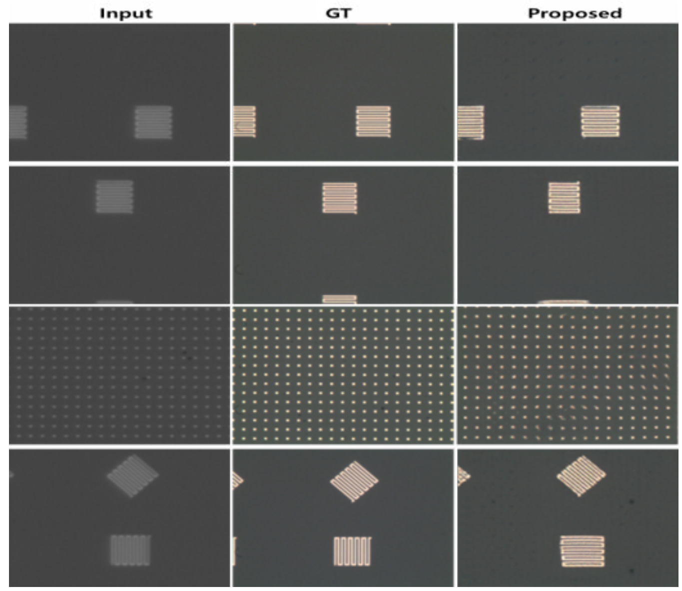


Fig. 3: The proposed method effectively generates high-magnification images from low-magnification inputs.

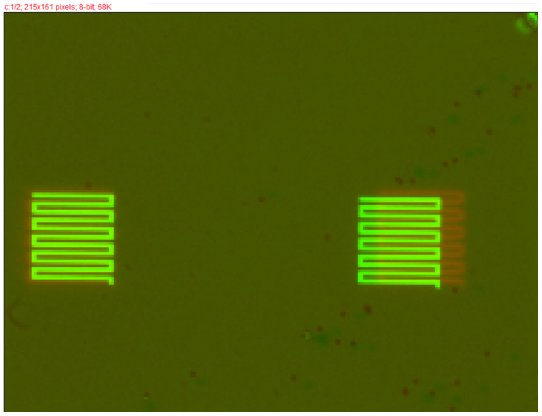
Across our test set, large, well-isolated defects are consistently generated with high fidelity, and defects located directly on top of mask patterns are generally well reproduced. In contrast, small defects in pattern-free regions tend to be omitted or appear blurred in the generated images. Since the ultimate goal of image synthesis is to facilitate defect detection—particularly in patterned regions where contrast is inherently low—the strength of our method lies in enhancing the visibility of pattern-overlapping defects. We therefore anticipate that combining the generated outputs for pattern-rich areas with the original imagery for pattern-free zones will yield the most robust overall defect detection performance.

V. Conclusion

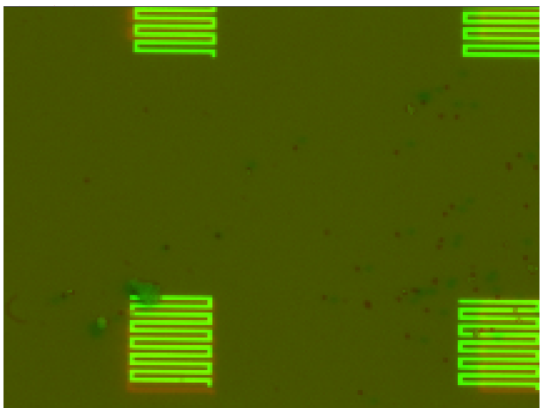
This paper presents the first application of a pix2pix-based generative model for synthesizing high-magnification extreme ultraviolet (EUV) mask images from low-magnification captures. We introduce a two-step matching pipeline that combines normalized cross-correlation template matching with edge-based contour refinement to precisely align low- and high-magnification image pairs, thereby producing high-quality training data for our pix2pix network. Experimental results show that our method faithfully synthesizes defects located on complex EUV mask patterns—defects that are nearly indistinguishable in the original low-magnification images—and that the generated high-magnification outputs make these defects readily detectable. This capability is expected to enable reliable identification of EUV pattern-overlapping defects that would otherwise evade detection under standard inspection conditions.

Acknowledgment

This work was supported by the Ministry of Trade, Industry & Energy(MOTIE, Korea), KIAT(Korea Institute for Ad-



(a) Error image 1



(b) Error image 2

Fig. 4: Representative error cases in generated EUV mask outputs: (a) One pattern is correctly generated, while the adjacent pattern is missing. (b) A pattern is generated with a 180° rotation error.

vancement of Technology) [No.P0023760 EUV Pellicle/Mask Database Construction and Defect Detection AI Solution for EUV Semiconductor Productivity Improvement] and the Technology Innovation Program(or Industrial Strategic Technology Development Program-Key Industry IT Convergence)(RS-2025-16932971, Development and demonstration of AI-based quality inspection technology for EUV mask defect inference and classification) funded By the Ministry of Trade Industry & Energy(MOTIE, Korea)

References

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems, vol. 27, 2014.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2414–2423.

[4] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[5] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in International Conference on Machine Learning, 2019.

[6] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4681–4690.

[7] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*, Munich, Germany, Sep. 2018, pp. 63–79.

[8] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 4172–4182.

[9] W. Podell, P. Esser, J. Rombach, A. Blattmann, B. Ommer, P. Dhariwal, and A. Brock, "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," arXiv preprint arXiv:2307.01952, Jul. 2023.

[10] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical Text-Conditional Image Generation with CLIP Latents," arXiv preprint arXiv:2204.06125, Apr. 2022.

[11] C. Schuhmann *et al.*, "LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models," *arXiv preprint arXiv:2210.08402*, Oct. 2022.

[12] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[13] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, Nov. 2014.

[14] Q. Zhu, K. Lu, M. Huo, and Y. Li, "Image-to-Image Translation with Diffusion Transformers and CLIP-Based Image Conditioning," *arXiv preprint arXiv:2505.16001*, May 2025.

[15] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, 2020.

[16] S. Chitwan, H. Jonathan, C. William, S. Tim, F. David, and N. Mohammad, "Image Super-Resolution via Iterative Refinement," *arXiv preprint arXiv:2104.07636*.

[17] G. Zihan, S. Haijian, D. Xing, and J. Yingtao, "DenoSR: A high-fidelity super-resolution approach for noisy images," ETRI Journal

[18] J. Zhang and X. Ma, "Fast diffraction model of lithography mask based on improved pixel-to-pixel generative adversarial network," *Optics Express*, vol. 31, no. 15, pp. 24437–24452, 2023.

[19] S. S. Mohammed and H. Gökalp Clarke, "Conditional image-to-image translation generative adversarial network (cGAN) for fabric defect data augmentation," *Neural Computing and Applications*, vol. 36, pp. 20231–20244, 2024.

[20] H. Zheng et al., "Phase defect characterization using generative adversarial networks for extreme ultraviolet lithography," *Applied Optics*, vol. 62, no. 5, pp. 1243–1252, 2023.