Diffusion versus GAN Synthesis for Wafer Map Defect Classification: A Preliminary Cross-Backbone Study on WM-811K

Ghulam Ashraf [1] **Wen-Tin Lee** [1]

Abstract

Synthetic data augmentation is the dominant strategy for class imbalance in wafer map defect classification on the WM-811K benchmark, with generative adversarial networks (GANs) and diffusion models reported in separate prior studies. To our knowledge, no work has compared these two families under a single controlled protocol on this benchmark. We present a cross-backbone evaluation of a two-stage hinge-loss GAN (TS-HingeGAN) and Stable Diffusion fine-tuned with class-conditioned LoRA (DB-SD-LoRA) on eight ImageNet-pretrained backbones. Holding split, seed, per-class synthetic budget, and evaluation set fixed across 24 cells (8 backbones $\times$ 3 conditions) under stratified 5-fold CV with real-only test partitions, we vary only the synthesis source and the backbone. DB-SD-LoRA attains pooled Clean-FID 51.0 versus 106.4 for TS-HingeGAN. Downstream gains are backbone-dependent: lightweight CNNs gain 2.5–3.6 Macro-F1 points from diffusion, while the strongest backbones (TinyViT-21M, ConvNeXtV2-T) gain only 0.6–0.8 points—a saturation pattern that prior single-backbone studies cannot reveal. Our best cell, TinyViT-21M + DB-SD-LoRA + ECOC-SVM, attains 0.9459 ± 0.0074 Macro-F1 under 5-fold CV, exceeding the matched-protocol 0.9399 of SCRBLAA-Net.

1. Introduction

Wafer maps encode die-level pass/fail outcomes across a circular silicon substrate; their spatial signatures reveal upstream process anomalies, making pattern classification a yield-critical task. The WM-811K benchmark of Wu et al. (2015) is the de facto reference. It exhibits severe class imbalance: of 811,457 wafers, only 25,519 carry one of eight defect labels, with class sizes spanning nearly two orders of magnitude (65:1 ratio). Direct training on this distribution under-allocates representational capacity to operationally informative rare classes. The literature has developed two principal synthesis families: GAN-based methods including AdaBalGAN (Wang et al., 2019), DCGAN-driven upsampling (Ebayyeh et al., 2022; Park & You, 2023), and the global-to-local design G2LGAN (Tsai & Wang, 2025); and, more recently, denoising diffusion probabilistic models (Wu et al., 2024; Li et al., 2024). Each family has been studied largely in isolation, and reported numbers are not directly comparable because train–test splits, class sets, and per-class synthetic budgets differ across studies.

We address this gap on a single, controlled axis. Holding the split protocol, random seed, per-class synthetic budget, training schedule, and evaluation set fixed, we vary only (i) the synthesizer—a two-stage hinge-loss GAN (TS-HingeGAN) or Stable Diffusion fine-tuned with class-conditioned LoRA adapters (DB-SD-LoRA)—and (ii) the classification backbone. Eight ImageNet-pretrained backbones spanning lightweight CNNs (MobileNetV2/V3-L, EfficientNet-B0/V2-S, DenseNet121), a modern ConvNet (ConvNeXtV2-Tiny), and Vision Transformers (TinyViT-21M, Swin-Tiny) (Dosovitskiy et al., 2021; Liu et al., 2021) are evaluated under Softmax and ECOC-SVM (Dietterich & Bakiri, 1995) classifier heads.

Contributions. (a) A controlled, backbone-wide head-to-head of GAN- and diffusion-based synthesis on WM-811K under identical hyperparameters, splits, and synthetic budgets across 24 cells, evaluated with stratified 5-fold CV as the primary protocol. (b) To our knowledge, the first reported application to WM-811K of Stable Diffusion fine-tuned with LoRA adapters, together with TinyViT-21M (Wu et al., 2022) and ConvNeXtV2-Tiny (Woo et al., 2023) backbones. (c) A headline result of TinyViT-21M + DB-SD-LoRA + ECOC-SVM at 0.9459 ± 0.0074 Macro-F1 under stratified 5-fold CV, exceeding the 0.9399 of SCRBLAA-Net (Kang et al., 2025) on the matched protocol, and 0.9233 Macro-F1 on the 9-class extension within 0.7 points of the 0.9301 of Tsai & Wang (2025). (d) The empirical observation that synthesis effectiveness decreases with backbone

This research was sponsored by the National Science and Technology Council in Taiwan under grant No. NSTC 113-2221-E-017-008-. [1]Department of Software Engineering and Management, National Kaohsiung Normal University, Kaohsiung, Taiwan. Correspondence to: Ghulam Ashraf <611378002@mail.nknu.edu.tw>, Wen-Tin Lee <wtlee@mail.nknu.edu.tw>.

capacity—a saturation pattern reproduced across protocols and not systematically explored in prior single-backbone studies.

2. Related Work

WM-811K classification. Early work used hand-crafted features with classical learners; after Wu et al. (2015) enabled CNN baselines, deep CNNs with class-imbalanced training (Saqlain et al., 2020) and lightweight networks (Yu et al., 2023) became standard. Vision transformers have only recently been evaluated as wafer-map backbones (Mohammad & Ryu, 2025).

GAN-based synthesis. AdaBalGAN (Wang et al., 2019) pairs conditional adversarial synthesis with an adaptive per-class controller; DCGAN-based upsampling has been combined with capsule classifiers (Ebayyeh et al., 2022) and tailored to extreme imbalance (Park & You, 2023). The global-to-local GAN of Tsai & Wang (2025) uses two-stage training (Stage 1: global pretrain on all classes pooled; Stage 2: per-class fine-tune); we adopt this two-stage schedule unchanged for our GAN baseline.

Diffusion synthesis. Wu et al. (2024) apply unconditional DDPM to die-level crops on proprietary data; Li et al. (2024) introduce an auxiliary-classifier DDPM trained from scratch on MIR-WM811K and MixedWM38. Both train from scratch. We instead start from a pretrained Stable Diffusion checkpoint (Rombach et al., 2022) and specialise per class via LoRA adapters (Hu et al., 2022) under the DreamBooth recipe (Ruiz et al., 2023)—an option not previously reported on WM-811K to our knowledge.

3. Method

Dataset and splits. We use the 25,519-wafer 8-class defect subset of WM-811K (Wu et al., 2015). Class counts: Edge-Ring 9,680, Edge-Loc 5,189, Center 4,294, Local 3,593, Scratch 1,193, Random 866, Donut 555, Near-full 149. Each wafer is resized to 224×224 with bilinear interpolation and the single grayscale channel tiled three times for ImageNet-pretrained backbones. Stratified 5-fold CV with seed 42 is our primary protocol: per fold, 15,311 train / 5,104 val / 5,104 test. Test partitions are real-only and identical across all synthesis conditions.

Pipeline. Following Figure 1, each cell (a) filters to eight defect classes, (b) applies the 5-fold split (seed 42), (c) optionally injects synthetic samples into training only, (d) fine-tunes a pretrained backbone with focal cross-entropy (Lin et al., 2017), (e) fits an ECOC-SVM head on penultimate features from the best-validation checkpoint, and (f) reports on real-only test. Steps (a)–(b) and (d)–(f) are held identical across all 24 cells; only the synthesis source in (c) varies.

TS-HingeGAN. Our GAN baseline pairs two-stage training (Tsai & Wang, 2025) with the hinge adversarial loss (Lim & Ye, 2017). The generator maps a 128-dim Gaussian latent to a $64 \times 64 \times 3$ output via a fully-connected layer to a $4 \times 4 \times 512$ volume and four nearest-neighbour $2\times$ upsample blocks. The discriminator is a five-layer convolutional stack with spectral normalisation (Miyato et al., 2018) on every layer. Hinge loss: $L_D = \mathbb{E}[\max(0, 1 - D(x))] + \mathbb{E}[\max(0, 1 + D(G(z)))]$; $L_G = -\mathbb{E}[D(G(z))]$, with a 2:1 G:D update ratio. Stage 1: one global generator on all classes pooled (3 epochs $\times$ 10,000 iter). Stage 2: per-class generators fine-tuned from this initialisation (10 epochs $\times$ 10,000 iter each). Adam, (β_1=0, β_2=0.9), $\mathrm{lr}_G = 1 \times 10^{-4}$, $\mathrm{lr}_D = 4 \times 10^{-4}$, batch 64.

DB-SD-LoRA. Our diffusion synthesizer uses Stable Diffusion v1.5 (Rombach et al., 2022) as a frozen latent diffusion backbone. Class specialisation is achieved through low-rank adapters (Hu et al., 2022) of rank r=32 (α=32) inserted into the `to_q/k/v/out` and `add_k/v_proj` projections of every UNet transformer block, trained under the DreamBooth recipe (Ruiz et al., 2023) with the standard noise-prediction MSE objective. A separate adapter is learned per defect class with a shared instance prompt; class identity is encoded by the adapter. Inference uses the UniPC scheduler at 100 denoising steps with guidance scale 7.5; outputs are downsampled $512 \rightarrow 64$. A CNN-based circular-shape filter rejects boundary-violating samples, reducing a raw pool of $\sim$75,000 to 27,927 retained samples and lowering pooled Clean-FID from 145.4 to 51.0.

Backbones and heads. Eight `timm` backbones: MobileNetV2 (Sandler et al., 2018), MobileNetV3-Large (Howard et al., 2019), EfficientNet-B0 (Tan & Le, 2019), EfficientNetV2-S (Tan & Le, 2021), DenseNet121 (Huang et al., 2017), TinyViT-21M (Wu et al., 2022), ConvNeXtV2-Tiny (Woo et al., 2023), Swin-Tiny (Liu et al., 2021). Each is fine-tuned with focal cross-entropy (γ=2.0) end-to-end for 30 epochs using AdamW (lr=1×10^{-4}, weight decay 5×10^{-2}) with cosine schedule and mixed precision. Online augmentation (horizontal/vertical flip, $\pm 15°$ rotation) is applied during classifier training. Best-validation accuracy selects the checkpoint. The ECOC-SVM head is fitted post-hoc on penultimate features over an RBF SVM (C=10, code size = 2.0) and evaluated on real-only test features.

Training protocol. Identical hyperparameters across all 24 cells. Synthetic samples are appended to the training partition only; validation and test partitions are real-only and locked by seed 42. Synthesised 64×64 samples are bilinearly upsampled to 224×224 before concatenation.

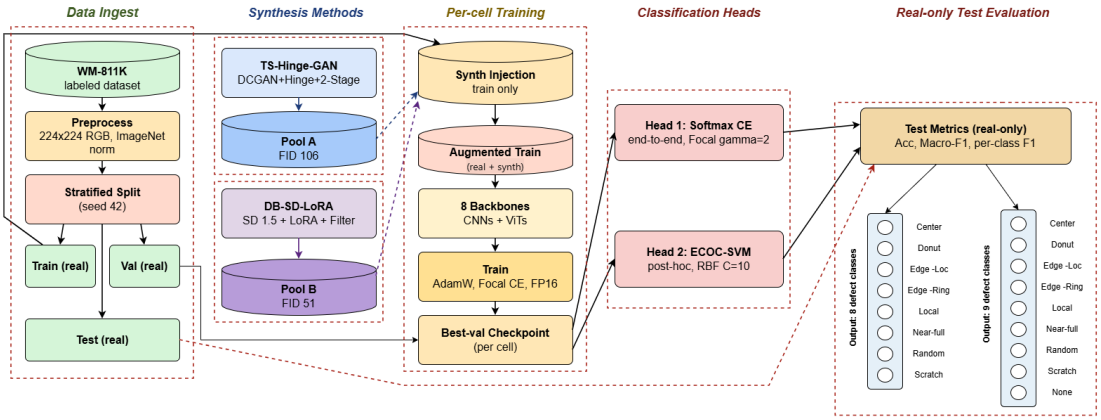

Figure 1. End-to-end pipeline. Across all 24 cells, only the synthesis source (none, TS-HingeGAN, or DB-SD-LoRA) varies; test partitions are real-only and locked across conditions.

The 5-fold protocol uses target 3× cap (per-class target 5,000, capped at 3× real); Edge-Ring synthetic count is zero (TS-HingeGAN failed to converge there). All cells are trained on a single NVIDIA RTX 5090 (32 GB VRAM) in PyTorch with mixed precision.

4. Results

4.1. Synthesis Quality

Pooled Clean-FID against the labelled real set favours diffusion by a factor of two: 51.0 for DB-SD-LoRA vs. 106.4 for TS-HingeGAN. DB-SD-LoRA attains sub-50 FID on Center, Edge-Ring, Local, Random, and Scratch; the two largest residuals—Donut (FID 92.6, size 555) and Near-full (FID 85.1, size 149)—are the two smallest training populations, where fine-tuning is sample-limited. TS-HingeGAN exceeds 100 FID on every class and failed to converge on Edge-Ring. The 27,927-sample retained pool from the circular-shape filter is used in every downstream cell.

TS-HingeGAN's non-convergence occurs on Edge-Ring, the largest defect class (9,680 wafers). We attribute this to instability of two-stage per-class fine-tuning when the per-class population is large and visually homogeneous, which collapses the Stage 2 adversarial signal. DB-SD-LoRA shows no such failure—reaching sub-50 FID on Edge-Ring—indicating that adapter-based specialisation of a pretrained diffusion prior is more robust to per-class population extremes, a qualitative advantage not captured by the pooled FID comparison.

4.2. Primary Results: 5-Fold Cross-Validation

Table 1 reports ECOC-SVM Macro-F1 (mean ± std over five folds, real-only test) for every backbone under each of the three conditions. The headline cell—TinyViT-21M + DB-SD-LoRA—reaches 0.9459 ± 0.0074; the same backbone with TS-HingeGAN reaches 0.9457 ± 0.0060 and with real data alone 0.9378 ± 0.0059.

Table 1. Primary results: ECOC-SVM Macro-F1 (mean ± std) under stratified 5-fold CV. Bold marks the best per row. +GAN = +TS-HingeGAN; +SD-LoRA = +DB-SD-LoRA. Standard deviations 0.003–0.010 across folds. Real-only cells use no online augmentation; synthesis cells use flip + $\pm 15^\circ$ rotation.

Backbone	Real	+GAN	+SD-LoRA
MobileNetV2	.8981	.9307	**.9344**
MobileNetV3-L	.9136	.9380	**.9382**
EfficientNet-B0	.9229	**.9410**	.9370
EfficientNetV2-S	.9228	**.9390**	.9380
DenseNet121	.9298	.9403	**.9420**
TinyViT-21M	.9378	.9457	**.9459**
ConvNeXtV2-T	.9329	**.9398**	.9390
Swin-Tiny	.9301	.9426	**.9432**

All eight backbones gain from at least one synthesis source, but the gain depends strongly on backbone capacity: 2.5–3.6 Macro-F1 points for lightweight CNNs (MobileNetV2/V3-L), 1.2–1.5 for mid-capacity backbones, and 0.6–0.8 for the strongest (TinyViT-21M, ConvNeXtV2-T)—a saturation regime that prior single-backbone studies cannot reveal. DB-SD-LoRA strictly exceeds TS-HingeGAN in 5 of 8 backbones (MobileNetV2, MobileNetV3-L, DenseNet121, TinyViT-21M, Swin-Tiny); two of the remaining three (EfficientNetV2-S, ConvNeXtV2-T) are within one fold's standard deviation. The same saturation pattern is reproduced under a supplementary 80/10/10 protocol.

4.3. Comparison with Prior Work

Table 2 positions our results against recent WM-811K studies. Our 8-class 5-fold Macro-F1 of 0.9459 matches/exceeds the 0.9399 of SCRBLAA-Net under the same protocol. On the 9-class 70/30 protocol of Tsai & Wang (2025), our Macro-F1 of 0.9233 is within 0.7 points of 0.9301. To per-

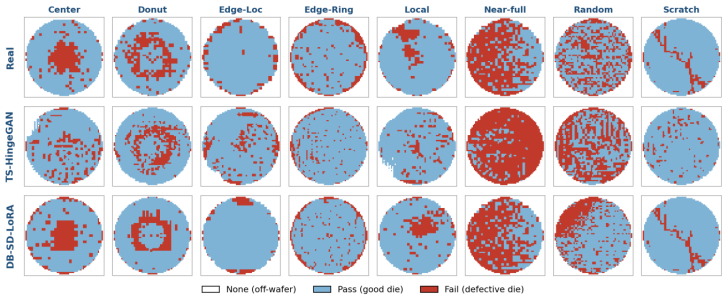

Figure 2. Sample grid: real (top), TS-HingeGAN (middle), and DB-SD-LoRA (bottom) across the eight defect classes. Diffusion preserves wafer boundary and defect signatures more faithfully than the GAN, particularly on Donut and Random.

Table 2. Comparison with prior work on WM-811K. Methods reported under their original protocols. The leaky-split row matches I-CBAM only.

Method	Split	F1	Acc
8-class protocol			
I-CBAM-ResNeXt50 (Chen et al., 2023a)	80/20 (leaky)	.9696	.9696
MFFP-Net (Chen et al., 2023b)	80/10/10	N/A	.9671
SCRBLAA-Net (Kang et al., 2025)	5-fold CV	.9399	.9498
Ours: TinyViT+SD-LoRA	5-fold CV	**.9459**	**.9623**
Ours: TinyViT+SD-LoRA	80/20 (leaky)	.9774	.9787
9-class protocol (with None)			
G2LGAN+MNv2 (Tsai & Wang, 2025)	70/30	.9301	.9839
Shin & Yoo (Shin & Yoo, 2023)	K-fold	.895	.980
Ours: TinyViT+SD-LoRA	70/30	.9233	.9842

mit controlled comparison with I-CBAM-ResNeXt50 (Chen et al., 2023a), whose split is applied after augmentation (test leakage), we include a matched leaky-split row; our best configuration reaches 0.9774 Macro-F1 under that protocol, indicating prior leaky-split numbers are not directly comparable to clean-split results.

5. Discussion

Three observations stand out. (i) The factor-of-two FID advantage of DB-SD-LoRA does not translate into a uniformly large downstream gain: the Macro-F1 gap between synthesizers is below 0.5 points in every backbone. Synthesis value depends more on backbone capacity to absorb the added distributional support than on synthesiser fidelity—lightweight CNNs gain most, the strongest least. (ii) The ECOC-SVM head improves over Softmax in 23 of 24 cells; being post-hoc (features from a fixed backbone, SVM fit once), it is a low-cost addition to any pipeline. (iii) EfficientNet-B0 + TS-HingeGAN (4.0 M, 0.9410) approaches the TinyViT headline within 0.5 points at one-fifth the cost.

On the value of small gains. Although the gain on the strongest backbones is small (0.6–0.8 points), its operational value is disproportionate: Macro-F1 weights all eight classes equally, so these gains fall on the rare defect classes (Donut, Near-full, Scratch)—the costly failure modes in inline inspection. At 10^5–10^6 wafers per run, a small rise in rare-class recall flags many more high-cost defects, so the deployment-relevant metric is tail recall, not average accuracy.

Generation overhead. The circular-shape filter rejects roughly 63% of raw generations ($75\text{K} \rightarrow 27.9\text{K}$). This is a one-time, *offline* cost incurred once per class before training; it adds nothing to classifier training or inference, and the retained pool is reused across all 24 cells, so it does not affect deployment cost.

Limitations. (i) single seed (multi-seed sweeps reserved for the journal extension); (ii) single dataset (cross-dataset validation on MIR-WM811K and MixedWM38 left to future work); and (iii) only off-the-shelf backbones, leaving custom wafer-map architectures open.

6. Conclusion

We presented a controlled cross-backbone comparison of GAN- and diffusion-based synthesis on WM-811K under stratified 5-fold CV. TinyViT-21M + DB-SD-LoRA + ECOC-SVM reaches 0.9459 Macro-F1 (8-class) and 0.9233 (9-class), matching or exceeding matched-protocol baselines, while showing that synthesis value is backbone-dependent and saturates with capacity—a pattern single-backbone studies cannot reveal.

References

Chen, S. et al. Wafer map defect pattern detection method based on improved attention mechanism. *Expert Systems with Applications*, 230:120544, 2023a.

Chen, Y. et al. Wafer defect recognition method based on multi-scale feature fusion. *Frontiers in Neuroscience*, 17:1202985, 2023b.

Dietterich, T. G. and Bakiri, G. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.

Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

Ebayyeh, A. A. R. M. A., Danishvar, S., and Mousavi, A. An improved capsule network (wafercaps) for wafer bin map classification based on dcgan data upsampling. *IEEE Transactions on Semiconductor Manufacturing*, 35(1):50–59, 2022.

Howard, A. et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, 2019.

Hu, E. J. et al. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Huang, G. et al. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.

Kang, G., Lee, G., and Kim, Y. Combining residual network and bidirectional long short-term memory with additive attention for wafer defect classification. *International Journal of Advanced Manufacturing Technology*, 141(9–10):4967–4983, 2025.

Li, J. et al. Sample-imbalanced wafer map defects classification based on auxiliary classifier denoising diffusion probability model. *Computers & Industrial Engineering*, 192:110209, 2024.

Lim, J. H. and Ye, J. C. Geometric gan. *arXiv preprint arXiv:1705.02894*, 2017.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.

Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Mohammad, F. and Ryu, D. Semiconductor wafer map defect classification with tiny vision transformers. *arXiv preprint arXiv:2504.02494*, 2025.

Park, S. and You, C. Deep convolutional generative adversarial networks-based data augmentation method for classifying class-imbalanced defect patterns in wafer bin map. *Applied Sciences*, 13(9):5507, 2023.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Ruiz, N. et al. Dreambooth: Fine-tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22500–22510, 2023.

Sandler, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520, 2018.

Saqlain, M., Abbas, Q., and Lee, J. Y. A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 33(3):436–444, 2020.

Shin, E. and Yoo, C. D. Efficient convolutional neural networks for semiconductor wafer bin map classification. *Sensors*, 23(4): 1926, 2023.

Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.

Tan, M. and Le, Q. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning (ICML)*, pp. 10096–10106, 2021.

Tsai, T.-H. and Wang, C.-Y. Wafer map defect classification using deep learning framework with data augmentation on imbalance datasets. *EURASIP Journal on Image and Video Processing*, 2025:6, 2025.

Wang, J., Yang, Z., Zhang, J., Zhang, Q., and Chien, W.-T. K. Adabalgan: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Transactions on Semiconductor Manufacturing*, 32(3): 310–319, 2019.

Woo, S. et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16133–16142, 2023.

Wu, K. et al. Tinyvit: Fast pretraining distillation for small vision transformers. In *European Conference on Computer Vision (ECCV)*, pp. 68–85, 2022.

Wu, M.-J., Jang, J.-S. R., and Chen, J.-L. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 28 (1):1–12, 2015.

Wu, P.-H. et al. Elevating wafer defect inspection with denoising diffusion probabilistic model. *Mathematics*, 12(20):3164, 2024.

Yu, N. et al. Wafer map defect patterns classification based on a lightweight network and data augmentation. *CAAI Transactions on Intelligence Technology*, 8(3):1029–1042, 2023.