UFSMatAD: A Unified Framework for Few-Shot Material Anomaly Detection Across Nanofiber SEM and Wafer Imaging

Shih-Chih Lin*
International Intercollegiate Ph.D. Program
National Tsing Hua University
Hsinchu City, Taiwan

Shang-Hong Lai†
Department of Computer Science
National Tsing Hua University
Hsinchu City, Taiwan

Abstract

Automated detection of nanoscale defects in materials imagery is challenging due to scarce labels, high morphological variability, and strict latency requirements in inline inspection. We present UFSMatAD, a unified, parameter-efficient framework for few-shot, multi-class anomaly detection in SEM and wafer AOI images. UFSMatAD replaces decoder feed-forward networks with Adapter Blocks configured with two bottleneck sizes and uses deterministic routing at inference to ensure stable optimization and predictable latency. A lightweight reconstruction head produces pixel-level maps and image-level scores. With the backbone frozen and only the decoder and head trainable, UFSMatAD matches transformer and diffusion baselines while substantially reducing trainable parameters and computational cost, and it remains robust under SEM-to-AOI domain shift. These results indicate that deterministic adapter mixtures provide a practical, scalable path to generalizable and resource-efficient industrial inspection.

1 Introduction

Automated defect detection is central to modern materials science and manufacturing, where microscopic irregularities degrade performance, reduce yield, and compromise safety in industrial and biomedical applications. Scanning Electron Microscopy (SEM) provides nanometer-scale resolution to capture defects, such as voids, beads, cracks, and films, in nanofibrous and composite materials [1]. Yet interpreting SEM images at scale remains difficult due to morphological variability, operator bias, and the high cost of expert annotations, motivating anomaly detection (AD) frameworks that can generalize across material types.

Traditional approaches relied on hand-crafted features to model normal structures and flag deviations [1]. Although sometimes effective on small datasets, they struggle with high intra-class variability and do not scale to production pipelines. Recent deep methods leverage CNNs and transformers for SEM classification and segmentation; dimensionality reduction (e.g., SVD/NMF) can improve efficiency and interpretability by reducing redundancy and revealing latent structure [2]. Large vision models tailored to materials, such as MatSAM [3], further suggest that domain-aware prompting can enable zero- or few-shot microstructure extraction in microscopy. However, three practical challenges persist: (i) data scarcity and imbalance—few labeled anomalies per class; (ii) generalization across categories and domains—defects vary widely in shape, scale, and texture, and

*Email: leolin65@gapp.nthu.edu.tw
†Email: lai@cs.nthu.edu.tw
Code: `https://github.com/leolin65/NIPS2025/tree/main`

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

deployment data often shift from development data; and (iii) efficiency and deployment—dense attention and large feed-forward blocks raise latency and memory, complicating inline inspection.

Many AD systems adopt a single-class paradigm that trains one model per defect type or product [4–8]. While straightforward, this approach is costly to maintain across many classes. Multi-class AD (MCAD) systems such as UniAD [9], HVQ-Trans [10], and DiAD [11] aim to unify categories using transformers or diffusion models, but their parameter and FLOP budgets can be high at deployment. Mixture-of-Experts variants (e.g., MoEAD [12]) reduce average compute by activating subsets of experts. Yet, stochastic routing can introduce training variance and unpredictable inference cost, which is undesirable in real-time inspection.

This work. We propose **UFSMatAD**, a unified, parameter-efficient framework for few-shot, multi-class AD in SEM and wafer AOI images. UFSMatAD replaces decoder feed-forward networks with *Adapter Blocks*, each of which uses exactly two bottleneck adapters. During training we allow soft competition between the two adapters; at inference we use deterministic top-1 routing to ensure stable behavior and predictable latency. A lightweight reconstruction head produces pixel-level maps and image-level scores. To minimize footprint, the backbone is frozen and only the decoder (with two adapters per block) and the head are trainable. In all experiments, we report both *trainable* and *overall* parameters, together with FLOPs, to make deployment costs explicit. Empirically, UFSMatAD achieves competitive accuracy relative to transformer and diffusion baselines while substantially reducing the trainable parameter budget and compute, and it maintains robustness under the SEM→AOI domain shift.

Position in the ecosystem. UFSMatAD complements domain-tailored segmentation with large models (e.g., MatSAM [3]) by providing a lightweight detector/segmenter for low-label, multi-class AD under deployment constraints [9–11]. By freezing the backbone and adapting only small adapter pathways with deterministic routing, UFSMatAD targets inline inspection scenarios with strict latency and memory ceilings.

Contributions.

- A unified few-shot MCAD framework for SEM and wafer AOI that *replaces* decoder FFNs with Adapter Blocks (two bottlenecks per block) and employs deterministic routing at inference for predictable latency.
- A lightweight reconstruction head for joint pixel- and image-level scoring; only the decoder and head are trainable, and we report *trainable* vs. *overall* parameters to clarify deployment costs.
- Comprehensive evaluation across nanofiber SEM and wafer-style AOI benchmarks showing competitive accuracy with substantially reduced trainable parameters and compute, and robustness under cross-domain deployment.

2 Related Work

Unsupervised anomaly detection (AD) is central to industrial inspection and materials science, where reliable detection of micro/nano-scale defects directly impacts yield and safety. In wafer manufacturing, targets span unpatterned defects (scratches, particles) and patterned defects (opens, shorts, line contamination); the field has transitioned from manual inspection to machine vision and deep learning to meet accuracy and throughput requirements [13, 14].

2.1 Sparse and Embedding-Based AD

Early SEM AD relied on local patch models and sparse dictionaries (e.g., nanofibers with beads/films), which worked on narrow domains but were brittle under morphology shifts [1]. Embedding-based AD leverages ImageNet features with nonparametric memory or density modeling, such as PatchCore and DifferNet [15, 16]. These methods perform well on natural-image benchmarks but remain sensitive to domain shifts between natural textures and SEM/wafer imagery, limiting transferability without domain adaptation.

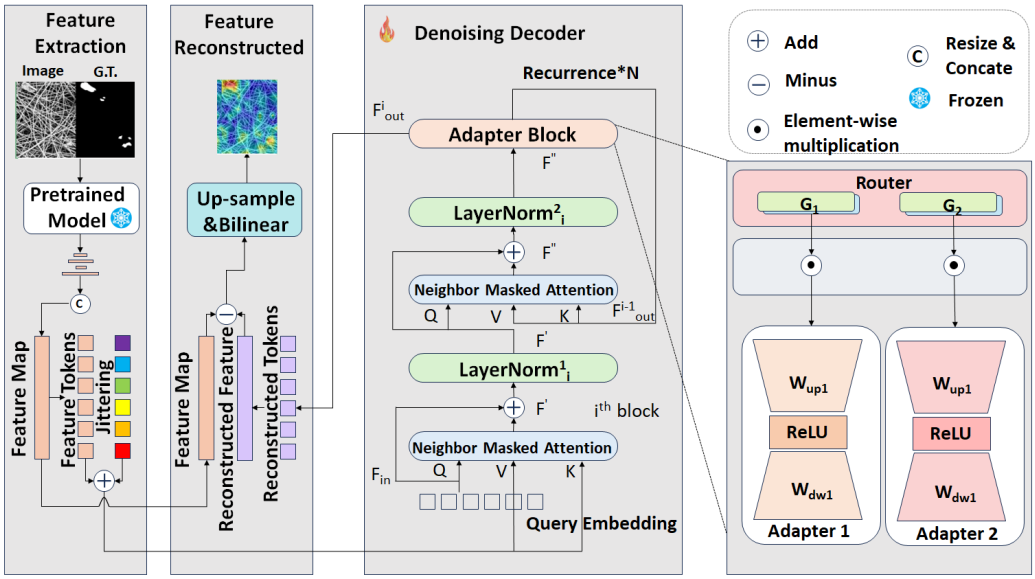

Figure 1: **Overview of UFSMatAD.** (1) A frozen backbone extracts multi-scale features. (2) A denoising decoder refines tokens by replacing FFNs with an adapters block; mild training-time feature jittering (spatial + Fourier) improves robustness. (3) Pixel-level anomalies are localized via reconstruction error and then bilinearly upsampled and Gaussian-smoothed. (4) Image-level scores combine peak and coverage statistics using a weighted max–mean–TopK aggregation.

2.2 Dimensionality Reduction in Deep Pipelines

SEM classification studies show that SVD/NMF can compress high-dimensional representations and improve efficiency while preserving discriminative power across CNN backbones [2]. Reported gains include reductions in speed and energy. However, these pipelines target classification and do not yield the pixel-level residual maps required by reconstruction-style AD.

2.3 Foundation Models and Prompting for Microscopy

Prompt-driven adaptations of SAM for materials microscopy (MatSAM) demonstrate broad generalization across OM/SEM datasets via shape-aware prompt fusion and domain-aware post-processing, rivaling or surpassing UNet/TransUNet on several segmentation tasks [3]. Yet such approaches are optimized for segmentation rather than anomaly detection, and their dense attention and prompt generation can hinder inline latency at scale.

2.4 Single-Class vs. Multi-Class AD

Synthesis- and reconstruction-based single-class pipelines (e.g., DRÆM and masked/diffusion variants) often require per-class training [4–8], and transformer/diffusion backbones introduce large FFNs and global attention that raise compute cost [10, 11]. Multi-class AD (MCAD) frameworks such as UniAD [9] unify categories with a single model, improving scalability but still inheriting heavy decoder blocks that stress real-time deployment.

2.5 MoE-Based Methods

MoEAD reduces compute by replacing decoder FFNs with sparsely activated experts for multi-class AD [12], but its stochastic routing and load-balancing can cause training instability and non-deterministic latency. We instead use lightweight Adapter Blocks with deterministic fusion (two bottleneck sizes by default), yielding stable optimization, predictable latency, and a small trainable footprint while supporting unified AD across SEM nanofibers and wafer-style AOI via a reconstruction head for pixel maps and an image-level score.

3 Proposed Method

3.1 Overview

We propose **UFSMatAD**, a parameter-efficient framework that replaces decoder feed-forward networks (FFNs) with adapter blocks and performs end-to-end, reconstruction-based anomaly scoring. We use lightweight adapters with *sparse, per-token* routing via a hard Gumbel–Softmax, yielding stable optimization and predictable latency—properties critical for inline SEM/AOI inspection. The model follows a compact backbone $\rightarrow$ denoising decoder $\rightarrow$ unified reconstruction–scoring pipeline, aligned with recent efficient unified AD designs that combine mild feature perturbations, neighbor-masked attention, and a reconstruction objective.

3.2 Stage 1: Multi-scale Feature Extraction

Given an input image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, a frozen backbone extracts spatial features:

$$\mathbf{X} \;=\; \mathrm{Backbone}(\mathbf{I}), \qquad \mathbf{X} \in \mathbb{R}^{C_{\mathrm{org}}\times H'\times W'}. \tag{1}$$

After resizing/concatenation, features are tokenized and projected to width D:

$$\mathbf{F}_{\mathrm{in}} \;=\; \mathrm{TokenizeProj}(\mathbf{X}), \qquad \mathbf{F}_{\mathrm{in}} \in \mathbb{R}^{L\times D},\;\; L = H'W'. \tag{2}$$

3.3 Stage 2: Denoising Decoder with Adapter block

Feature perturbations. To enhance robustness across devices and domains (e.g., SEM$\rightarrow$AOI), we inject mild perturbations before decoding:

$$\tilde{\mathbf{F}}_{\mathrm{in}} \;=\; \mathbf{F}_{\mathrm{in}} \;+\; \epsilon_s \;+\; \mathcal{F}^{-1}\big(\epsilon_f \odot \mathcal{F}(\mathbf{F}_{\mathrm{in}})\big), \tag{3}$$

where $\epsilon_s \sim \mathcal{N}(0,\sigma^2)$ (spatial noise) and ϵ_f jitters high-frequency bands in the Fourier domain; $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the (inverse) Fourier transform and $\odot$ is the Hadamard product.

Neighbor-Masked Attention (NMA). Inspired by UniAD [9], we use NMA, in which a learnable query attends only to local neighborhoods, thereby reducing leakage from anomalous regions. We mask a fixed $h_m \times w_m$ window on the $h \times w$ token grid around each query (a small window in all experiments), suppressing identity copying while preserving nonlocal context:

$$\mathbf{F}^{(1)} = \mathrm{NMA}\big(\mathbf{q},\, \tilde{\mathbf{F}}_{\mathrm{in}},\, \tilde{\mathbf{F}}_{\mathrm{in}}\big), \tag{4}$$

$$\mathbf{F}^{(2)} = \mathrm{LN}\big(\mathbf{F}^{(1)} + \tilde{\mathbf{F}}_{\mathrm{in}}\big), \tag{5}$$

$$\mathbf{F}^{(3)} = \mathrm{NMA}\big(\mathbf{F}^{(2)},\, \mathbf{F}^{\star},\, \mathbf{F}^{\star}\big), \tag{6}$$

$$\mathbf{F}^{(4)} = \mathrm{LN}\big(\mathbf{F}^{(3)} + \mathbf{F}^{(2)}\big). \tag{7}$$

Here $\mathbf{F}^{\star}$ is the previous block output; attention and adapter weights are shared across blocks to keep the decoder compact, with per-block LayerNorms.

Adapter block. Each adapter is a normalized bottleneck with LayerNorm at the input and output and a learnable scale s_k; by default, we use two bottleneck widths $d \in \{16, 32\}$ to trade coarse vs. fine refinements with negligible compute overhead:

$$\mathbf{U} = \mathrm{LN}\Big(\mathbf{F}^{(4)}\Big), \tag{8}$$

$$\mathbf{h}_k = \mathrm{GELU}\Big(\mathbf{W}_{\downarrow}^{(k)}\mathbf{U}\Big), \qquad \mathbf{W}_{\downarrow}^{(k)} \in \mathbb{R}^{D\times d_k}, \tag{9}$$

$$\mathbf{y}_k = \mathbf{W}_{\uparrow}^{(k)}\,\mathrm{Dropout}(\mathbf{h}_k), \qquad \mathbf{W}_{\uparrow}^{(k)} \in \mathbb{R}^{d_k\times D}, \tag{10}$$

$$A_k\Big(\mathbf{F}^{(4)}\Big) = s_k\,\mathrm{LN}(\mathbf{y}_k). \tag{11}$$

Sparse routing via Gumbel–Top1 (training). A linear gate produces per-token adapter logits; a hard Gumbel–Softmax selects exactly one adapter per token:

$$\mathbf{G} = \mathbf{F}^{(4)}\mathbf{W}_r \in \mathbb{R}^{L\times K}, \tag{12}$$

$$\mathbf{Z} = \mathrm{GumbelSoftmax}\big(\mathbf{G}/\tau,\ \text{hard} = \text{True}\big) \in \{0,1\}^{L\times K}, \tag{13}$$

$$A_{\mathrm{AdapterBlock}}\big(\mathbf{F}^{(4)}\big) = \big[\, A_{k^\star(1)}(\mathbf{F}_1^{(4)}) \ \cdots\ A_{k^\star(L)}(\mathbf{F}_L^{(4)}) \,\big]^\top, \tag{14}$$

where $k^\star(i) = \arg\max_k \mathbf{Z}_{i,k}$ and gradients use the straight-through estimator. The decoder block output is

$$\mathbf{F}_{\mathrm{dec}} \ = \ \mathbf{F}^{(4)} \ + \ A_{\mathrm{AdapterBlock}}\big(\mathbf{F}^{(4)}\big). \tag{15}$$

3.4 Unified Reconstruction and Scoring

Reconstruction head. Decoder outputs are projected back to the backbone channel space and reshaped:

$$\hat{\mathbf{F}} \ = \ \mathrm{OutProj}\big(\mathbf{F}_{\mathrm{dec}}\big) \ \in \ \mathbb{R}^{C_{\mathrm{org}}\times H'\times W'}. \tag{16}$$

Training objective: MSE + cosine dissimilarity. Let $\langle\cdot,\cdot\rangle$ denote the channel-wise inner product and $\|\cdot\|_2$ the ℓ_2 norm along channels. We minimize

$$\mathcal{L}_{\mathrm{rec}} \ = \ \underbrace{\big\|\mathbf{X}-\hat{\mathbf{F}}\big\|_2^2}_{\text{MSE}} + \alpha\Big(1-\cos\big(\mathbf{X},\hat{\mathbf{F}}\big)\Big), \qquad \cos(\mathbf{A},\mathbf{B}) \ = \ \frac{\langle\mathbf{A},\mathbf{B}\rangle}{\|\mathbf{A}\|_2\,\|\mathbf{B}\|_2+\varepsilon}, \tag{17}$$

where $\alpha > 0$ balances spatial fidelity and semantic alignment; averages are taken over spatial locations and the batch.

Pixel map and post-processing. We compute a pixel-level anomaly map using only reconstruction magnitude:

$$\mathbf{S}_{\mathrm{pix}} \ = \ \sqrt{\big\|\mathbf{X}-\hat{\mathbf{F}}\big\|_2^2}. \tag{18}$$

The map is upsampled to input resolution via bilinear interpolation and lightly Gaussian-smoothed to stabilize AUROC thresholds.

Image-level score. We aggregate peak and coverage statistics with fixed convex weights:

$$S_{\mathrm{img}} \ = \ w_1\ \max(\mathbf{S}_{\mathrm{pix}}) \ + \ w_2\,\mathrm{mean}(\mathbf{S}_{\mathrm{pix}}) \ + \ w_3\,\mathrm{TopKMean}(\mathbf{S}_{\mathrm{pix}}), \qquad \sum_{i=1}^{3} w_i = 1\,. \tag{19}$$

Training and inference. Training. We train on *normal* images only. Inputs are tokenized and projected to $\mathbf{F}_{\mathrm{in}}$, then mildly perturbed with spatial Gaussian noise and Fourier-band jitter. The denoising decoder (NMA + Adapter Block with Gumbel–Top1 routing; Sec. 3.3) processes these perturbed tokens and produces reconstructions. Parameters are optimized end to end using the reconstruction objective $\mathcal{L}_{\mathrm{rec}}$ (Eq. 17).

Inference. At test time, we disable all perturbations and run the same model deterministically on both *normal* and *anomalous* images. We report the pixel-wise anomaly map $\mathbf{S}_{\mathrm{pix}}$ (Eq. 18) and the image-level score S_{img} (Eq. 19).

4 Experiments

4.1 Benchmarks and Metrics

We evaluate **UFSMatAD** on (i) **Nanofiber SEM Defect** [1], a high-resolution SEM set with beads/voids/film-like defects under scarce labels and large morphology variance, and (ii) **Wafer AOI (Texture-AD)** [17], wafer-style texture anomalies with pixel-level masks across 14 subclasses and >14k samples. We report AUROC (%) at *image* (AUROCi) and *pixel* (AUROCp) levels; anomaly maps come from decoder reconstructions, bilinearly upsampled and Gaussian-smoothed ($\sigma \approx 1.0$) unless noted. Unless noted, training uses normal-only images; few-shot refers to the number of normal support samples per category.

Table 1: Few-shot anomaly detection on NanoFiber SEM [1] and Wafer AOI [17]. We report AUROC (%) at the image level (AUROCi) and pixel level (AUROCp) under 1/2/4-shot settings.

Dataset	**Metric**	**Few-1**	**Few-2**	**Few-4**
NanoFiber SEM	AUROCi	74.20	76.40	77.00
	AUROCp	74.90	90.00	97.50
Wafer AOI	AUROCi	74.20	76.40	80.00
	AUROCp	80.10	82.30	85.40

Table 2: Results on Nanofiber SEM Defect [1] (4-shots) and Wafer AOI [17] (4-shots). We report AUROC (%) at image level (AUROCi) and pixel level (AUROCp). The last column is the average of the four AUROC numbers. In tables, the best score is **bold** and the second best is underlined.

Method	NanoFiber SEM (few-shot) AUROCi	NanoFiber SEM (few-shot) AUROCp	Wafer AOI (few-shot) AUROCi	Wafer AOI (few-shot) AUROCp	Avg. AUROC
UniAD [9]	75.0	96.3	78.8	84.8	83.7
DiAD [11]	74.6	95.6	79.0	85.0	83.6
MoEAD [12]	76.7	97.2	79.8	85.2	84.7
UFSMatAD (ours)	**77.0**	**97.5**	**80.0**	**85.4**	**85.0**

4.2 Training Protocol and Implementation

All images are resized to 224×224 and normalized with ImageNet statistics. The encoder backbone is *frozen*; only the decoder and reconstruction head are trainable, and our adapter block replaces every Transformer FFN. The decoder has four layers, eight heads, and a hidden size of 256; Neighbor-Masked Attention (NMA) is used in both self- and cross-attention (Sec. 3.3). Attention and adapter weights are shared across layers to keep the decoder compact; LayerNorms are per-block. **Adapter configuration**: two bottlenecks ($d\in\{16, 32\}$). The `AdapterRouter` is trained with *Gumbel-Top1*; at inference, we use *deterministic routing* (no perturbations) for predictable latency. **Optimization**: AdamW (lr 2×10^{-4}, weight decay 10^{-4}), StepLR with $\gamma=0.1$ every 800 epochs, gradient clipping (max-norm 0.1), up to 500 epochs with validation every 25 epochs, batch size 128 on a single RTX 4090. **Objective**: hybrid reconstruction

$$\mathcal{L} = \|\mathbf{F}_{\text{rec}} - \mathbf{F}_{\text{align}}\|_2^2 + \alpha\big(1 - \cos(\mathbf{F}_{\text{rec}}, \mathbf{F}_{\text{align}})\big), \quad \alpha = 0.5.$$

Robustness augmentation: Gaussian noise in the spatial domain and Fourier high-frequency jittering ($q=0.3$) during training only; both are disabled at test time.

4.3 Experiments Results

Evaluation protocol. We compare UFSMatAD to embedding-, reconstruction-, diffusion-, and MoE-based AD baselines. All results report AUROC (%) at the *image level* (AUROCi) and *pixel level* (AUROCp), averaged over three seeds. In tables, the best score is **bold** and the second best is underlined.

Few-shot behavior across domains. Table 1 summarizes 1/2/4-shot performance on NanoFiber SEM [1] and Wafer AOI [17]. On **NanoFiber SEM**, UFSMatAD already performs well with a single support (few-1: 74.2 AUROCi / 74.9 AUROCp). Adding one more support (few-2) yields a large localization gain (AUROCp 74.9 $\rightarrow$ 90.0) with a modest AUROCi increase, indicating that extra spatial cues mainly help pixel-wise mapping. With few-4, AUROCp reaches 97.5, suggesting near-saturation under very limited supervision. On **Wafer AOI**, results are competitive at few-1 (74.2 AUROCi / 80.1 AUROCp) and improve steadily as shots increase (AUROCp 80.1 $\rightarrow$ 85.4 from few-1 to few-4). Compared to NanoFiber SEM, wafer gains are more gradual, consistent with structured grid-like patterns where improvements accrue with additional exemplars. Overall, UFSMatAD maintains stable image-level detection while scaling pixel-level localization effectively as shots increase.

Table 3: Compute vs. accuracy with *trainable* parameters only (frozen EfficientNet backbone). FLOPs are per forward pass at 224^2. Average AUROC is the mean of {NanoFiber SEM AUROCi, NanoFiber SEM AUROCp, Wafer AUROCi, Wafer AUROCp} from Table 2.

Method	Parameters (M)	FLOPs (G)	Average AUROC (%)
UniAD [9]	7.7	4.30	83.7
DiAD [11]	1300	$>$2200	83.6
MoEAD [12]	4.9	2.18	84.7
UFSMatAD (ours)	**1.3**	**1.90**	**85.0**

Notes: Counts report *trainable* decoder/head only; backbones are frozen. UFSMatAD uses two adapters (d=16, 32), hidden 256; trainable $\approx$1.3M; forward FLOPs $\approx$1.9G. DiAD's diffusion transformer is substantially heavier; literature reports $\sim$1.3B trainable params and $>$2.2T FLOPs.

Table 4: Adapter width ablation (4-shot). Trainable counts only (frozen backbone). FLOPs are per forward pass at 224^2.

			NanoFiber SEM (few-shot)		Wafer AOI (few-shot)		Avg. AUROC (%)
Adapters (d)	**Params (M)**	**FLOPs (G)**	**AUROCi**	**AUROCp**	**AUROCi**	**AUROCp**	
$\{16, 32\}$ (default)	**1.30**	**1.90**	77.0	97.5	80.0	85.4	**85.0**
$\{16, 32, 64, 128\}$	1.42	1.93	77.2	97.6	80.2	85.6	85.2

Benchmark comparison (4-shot). Table 2 compares 4-shot results on both datasets. UFSMatAD attains the best score on all four metrics and the cross-dataset average. On NanoFiber SEM, it improves image AUROC over UniAD by $+2.0$ (77.0 vs. 75.0) and edges out the strongest baseline on pixel AUROC by $+0.3$ (97.5 vs. 97.2 for MoEAD). On Wafer AOI, it provides consistent gains of $+0.2$ in both image AUROC (80.0 vs. 79.8) and pixel AUROC (85.4 vs. 85.2) relative to MoEAD. The mean AUROC reaches **85.0**, exceeding MoEAD (84.7), UniAD (83.7), and DiAD (83.6), while using a frozen backbone with only the decoder and head trainable ($\approx$1.3M params) and deterministic routing for predictable latency.

4.4 Computational Efficiency and Reproducibility

Trainable scope. Unless stated otherwise, the EfficientNet backbone is *frozen*; counts in Table 3 refer to *trainable* decoder+head parameters. UFSMatAD trains $\approx$1.3M parameters with two adapters (d=16, 32, hidden 256), compared to 4.9M (MoEAD), 7.7M (UniAD), and $\sim$1,300M (DiAD).

Compute. FLOPs are measured per forward pass at 224^2. UFSMatAD requires $\approx$1.90 G vs. 2.18 G (MoEAD), 4.30 G (UniAD), and $>$ 2,200 G (DiAD). For UFSMatAD, forward+backward is $\approx$5.7 G per image under the default two-adapter setting.

Determinism. At test time we disable perturbations and use deterministic routing (no stochastic gates), so inference is deterministic. All metrics are the mean over three training seeds affecting only initialization and data shuffling.

Reproducibility package. We release code and exact configs (YAML), dataset versions/splits, and full hyperparameters; hardware and counting rules are documented (single RTX 4090; frozen backbone; trainable counts; measured FLOPs).

4.5 Adapter Width Ablation

We compare two adapter configurations under the 4-shot protocol: a lightweight $\{16, 32\}$ setting versus an expanded $\{16, 32, 64, 128\}$ variant. The larger setting increases *trainable* parameters by $\sim$9% (1.30$\rightarrow$1.42 M) and forward FLOPs by $\sim$1.6% (1.90$\rightarrow$1.93 G), yielding only a **+0.2** average AUROC gain (Table 4). To prioritize efficiency and predictable latency, we adopt $\{16, 32\}$ for all main results.

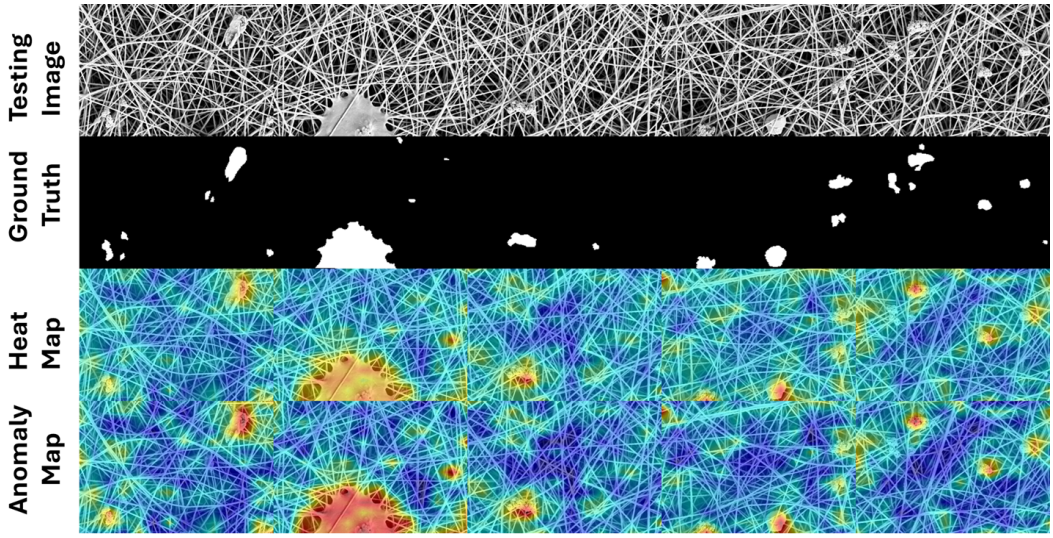

Figure 2: NanoFiber SEM examples. Top: input; middle: ground-truth mask; bottom: UFSMatAD anomaly map (pixelwise score from the decoder reconstruction), bilinearly upsampled, lightly Gaussian-smoothed, normalized to $[0, 1]$, shown as blue→red (red = more anomalous).

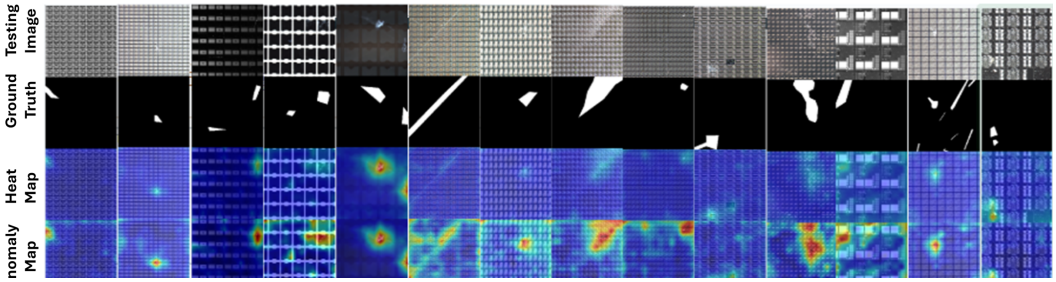

Figure 3: Wafer AOI visualization (**14 classes**). Top: input; middle: ground-truth mask; bottom: UFSMatAD anomaly map (pixelwise score from the decoder reconstruction), bilinearly upsampled, lightly Gaussian-smoothed, normalized to $[0, 1]$, blue→red (red = more anomalous).

5 Visualization

We provide qualitative results on **NanoFiber SEM** and **Wafer AOI**. Each triplet shows the input image, the ground-truth mask, and the predicted anomaly heatmap produced by our reconstruction head with the same light post-processing used in evaluation (bilinear upsampling and Gaussian smoothing). As illustrated in Fig. 2, UFSMatAD yields sharp, well-localized heatmaps that align with annotated defects, capturing both localized nanoscale irregularities (e.g., broken fibers, voids) and wafer-scale process defects (e.g., line scratches, pattern misalignments) without excessive activations on repetitive backgrounds. This behavior is consistent with our decoder design: *neighbor-masked attention* (NMA) constrains attention to local neighborhoods, and the *deterministic* Mixture-of-Adapters fusion used at inference (Sec. 3.3), together promoting stable localization under few-shot supervision. We also include typical failure cases in the supplement: (i) very faint, line-like scratches under strong illumination drift and (ii) extremely low-SNR SEM regions, where errors manifest as slightly diffuse maps rather than missed detections. All visualized heatmaps use the same bilinear upsampling and light Gaussian smoothing as in evaluation for consistency.

6 Conclusion

We presented **UFSMatAD**, a unified and parameter-efficient framework for few-shot, multi-class anomaly detection across SEM nanofibers and wafer-style AOI. UFSMatAD replaces decoder FFNs with a *deterministic* Mixture-of-Adapters and employs NMA in a compact denoising decoder, yielding predictable latency and stable optimization with a *frozen backbone* (trainable parameters $\approx$1.3M). Experiments show competitive or state-of-the-art AUROC at both image and pixel levels while using substantially fewer trainable parameters and FLOPs than recent unified AD baselines. Results further indicate distinct adaptation profiles across irregular SEM textures and structured wafer grids, consistent with our lightweight reconstruction scoring and mild, *training-time* robustness perturbations.

Limitations and future work. UFSMatAD can under-emphasize faint subpixel line defects under illumination drift, produce slightly diffuse maps in very low-SNR SEM regions, and degrade under domain shift or rare morphologies. Future work includes frequency-aware priors with self-supervised denoising, uncertainty calibration, temporal consistency for SEM video, multimodal fusion with process metadata/spectroscopy (e.g., EBSD/EDS), and broader cross-domain tests with active sampling. Beyond 2D, we will extend to materials 3D datasets (micro-CT, XRM, FIB-SEM) via volumetric backbones and voxel-level metrics, enforcing slice consistency and handling anisotropic spacing with memory-efficient patch inference and sparse attention.

References

[1] Diego Carrera, Fabio Manganini, Giacomo Boracchi, and Ettore Lanzarone. Defect detection in sem images of nanofibrous materials. *IEEE Transactions on Industrial Informatics*, 13(2): 551–561, 2016.

[2] Cagri Yardimci and Mevlut Ersoy. Optimizing deep learning architectures for sem image classification using advanced dimensionality reduction techniques. *Research on Engineering Structures and Materials*, 11, 2025.

[3] Changtai Li, Xu Han, Chao Yao, and Xiaojuan Ban. Matsam: Efficient materials microstructure extraction via visual large model. *CoRR*, 2024.

[4] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021.

[5] Shih Chih Lin, Ho-Weng Lee, Yu-Shuan Hsieh, Cheng Yu Ho, and Shang-Hong Lai. Masked attention convnext unet with multi-synthesis dynamic weighting for anomaly detection and localization. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2023.

[6] Shih-Chih Lin and Shang-Hong Lai. Clip-fsqae: Clip-guided finite scalar quantized autoencoder for few-shot anomaly detection. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2025.

[7] Shih-Chih Lin and Shang-Hong Lai. LFQUIAD: Lookup-free quantized autoencoder for few-shot unsupervised industrial anomaly detection via synthetic diffusion inpainting. Presented at the Synthetic Data for Computer Vision Workshop, CVPR, 2025. `https://syndata4cv.github.io/`.

[8] Shih-Chih Lin and Shang-Hong Lai. Squad: Scalar quantized representation learning for unsupervised anomaly detection and localization. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2024.

[9] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022.

[10] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 36:8487–8500, 2023.

[11] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 8472–8480, 2024.

[12] Shiyuan Meng, Wenchao Meng, Qihang Zhou, Shizhong Li, Weiye Hou, and Shibo He. Moead: A parameter-efficient model for multi-class anomaly detection. In *European Conference on Computer Vision*, pages 345–361. Springer, 2024.

[13] Jianhong Ma, Tao Zhang, Cong Yang, Yangjie Cao, Lipeng Xie, Hui Tian, and Xuexiang Li. Review of wafer surface defect detection methods. *Electronics*, 12(8):1787, 2023.

[14] Shih-Chih Lin and Shang-Hong Lai. Unilead: A unified and lightweight model for anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1472–1480, 2025.

[15] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022.

[16] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but differnet: Semi-supervised defect detection with normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1907–1916, 2021.

[17] Tianwu Lei, Bohan Wang, Silin Chen, Shurong Cao, and Ningmu Zou. Texture-ad: An anomaly detection dataset and benchmark for real algorithm development. *arXiv preprint arXiv:2409.06367*, 2024.