ISSN: 3079-6342

# Transformer Networks Enable Nanoscale Defect Detection at Advanced Semiconductor Nodes

Zhixuan Gao*, Haoyu Deng, and Ethan Marshall

Department of Electrical and Computer Engineering, Texas Tech University, USA

*** Corresponding author:** zhixuan2658@gmail.com

## Abstract

**The relentless scaling of semiconductor manufacturing toward sub-5nm process nodes has rendered traditional optical and rule-based inspection systems increasingly inadequate for detecting nanoscale structural anomalies. This paper proposes a transformer-based deep learning framework for automated nanoscale defect detection in advanced semiconductor fabrication environments. The proposed architecture leverages self-attention mechanisms and multi-head feature extraction to capture long-range spatial dependencies across scanning electron microscopy (SEM) images of wafer surfaces, enabling high-sensitivity identification of defects including voids, bridging, line-edge roughness, and particle contamination. A hybrid pipeline combining a convolutional feature extractor with a Vision Transformer (ViT) encoder is developed and evaluated on a multi-class wafer defect dataset. Experimental results demonstrate that the proposed model achieves a defect detection accuracy of 97.3%, a false negative rate of 1.8%, and a mean average precision of 95.6%, outperforming conventional convolutional neural network (CNN)-based and support vector machine (SVM)-based baselines. The findings confirm that transformer architectures offer a compelling pathway for scalable, high-throughput semiconductor quality control, particularly as device geometries continue to shrink toward physical limits.**

## Keywords

**transformer networks, semiconductor defect detection, Vision Transformer, scanning electron microscopy, wafer inspection, self-attention, nanoscale manufacturing**

## 1. Introduction

The semiconductor industry has entered an era of extraordinary manufacturing complexity. As integrated circuit (IC) device geometries scale below 5 nanometers under leading-edge process nodes, the physical tolerances governing pattern fidelity, layer alignment, and material uniformity become vanishingly small. A single defect at this scale — whether a void in a metal interconnect, an unintended bridge between adjacent fins, or a particle deposited during deposition — can propagate through hundreds of downstream process steps before manifesting as a yield-limiting failure. The economic consequences of undetected defects are therefore enormous: yield loss at advanced nodes can account for tens of millions of dollars per quarter for a major fab, and the latency between defect introduction and detection remains a critical bottleneck in modern semiconductor manufacturing control [1]. Historically, defect detection has relied on optical inspection systems augmented by rule-based classification algorithms. These approaches, while effective at micrometer-scale feature sizes, face fundamental physical limitations at sub-10nm geometries. Optical diffraction limits constrain the spatial resolution achievable with broadband illumination, and the complexity of modern three-dimensional transistor architectures — including FinFET and gate-all-around (GAA) structures — introduces defect modes that do not manifest clearly in two-

ISSN: 3079-6342

dimensional optical images [2]. As a result, the semiconductor industry has increasingly turned to electron beam inspection and SEM as primary metrology modalities for nanoscale defect characterization. SEM delivers sub-nanometer imaging resolution capable of resolving critical-dimension (CD) variations and surface topography at the feature level, but it also generates image datasets of enormous volume and complexity that overwhelm manual review workflows [3]. The emergence of deep learning has transformed automated image analysis across many domains, and semiconductor inspection is no exception. CNN architectures have been widely applied to wafer defect classification, demonstrating strong performance on pattern-matching tasks when sufficient labeled training data is available [4]. However, CNN architectures are inherently constrained by the locality of their receptive fields: each convolutional filter aggregates spatial information only within a fixed kernel window, limiting the model's ability to capture global structural relationships that are often diagnostic of certain defect classes. For instance, systematic defects resulting from process drift — such as periodic line-width variations across an exposure field — may span hundreds of pixels in an SEM image, yet a standard CNN with a small kernel would process them as collections of local texture features rather than as coherent global patterns [5]. A critical architectural development that partially addressed this locality limitation was the introduction of multi-scale feature pyramid representations. As illustrated conceptually in approaches like the Feature Pyramid Network (FPN), detection architectures evolved from single-resolution processing toward hierarchical multi-scale feature construction, enabling models to simultaneously reason about fine-grained local features and coarser global structural context [6]. This multi-scale philosophy proved particularly valuable for defect detection tasks where anomalies manifest across a wide range of spatial scales, from sub-pixel voids to multi-feature-spanning bridging patterns. However, even with multi-scale feature pyramids, the fundamental locality constraint of convolutional processing remained a barrier to modeling the long-range spatial correlations characteristic of systematic process-induced defects. Transformer architectures, originally developed for natural language processing (NLP) tasks, address this limitation through the self-attention mechanism, which enables every element in a sequence to attend to every other element simultaneously, regardless of positional distance. The adaptation of transformers to image recognition tasks — most prominently through the ViT framework — has opened a new chapter in computer vision, enabling models to reason about global spatial context with a degree of flexibility unavailable to purely convolutional approaches [7]. In the context of semiconductor defect detection, this capability is particularly valuable: defect signatures often depend not only on local pixel patterns but on their spatial relationship to surrounding structures, neighboring features, and global layout topology. This paper presents a transformer-based defect detection framework specifically designed for nanoscale semiconductor inspection. The system integrates a lightweight convolutional backbone for initial feature map construction with a multi-layer transformer encoder that models global spatial dependencies, followed by a classification head for multi-class defect categorization. The architecture is evaluated against established CNN and SVM baselines on a diverse wafer defect dataset spanning eight defect categories. The contributions of this work include a novel hybrid architecture tailored to semiconductor inspection constraints, a systematic ablation study isolating the contribution of the attention mechanism, and a performance benchmarking analysis demonstrating significant improvements in detection accuracy and false negative rate (FNR) reduction — both critical metrics for high-volume manufacturing (HVM) deployment.

## 2. Literature Review

The application of machine learning to semiconductor wafer defect inspection has evolved substantially over the past decade, transitioning from handcrafted feature engineering to end-

ISSN: 3079-6342

to-end learned representations. Early automated inspection systems relied heavily on statistical process control (SPC) frameworks combined with classical image processing techniques such as morphological filtering, Gabor wavelets, and principal component analysis (PCA) to characterize defect signatures on wafer maps [8]. While these methods offered interpretability and low computational overhead, their performance degraded significantly when applied to novel defect types not captured in the original feature design, limiting their utility in dynamic manufacturing environments where new process-induced defect modes emerge regularly. The adoption of CNNs for wafer defect classification marked a significant inflection point. Nakazawa and Kulkarni demonstrated that deep CNNs trained on wafer map images could achieve classification accuracies exceeding 98% across standard defect pattern categories including scratch, ring, and local cluster patterns, substantially outperforming traditional machine learning approaches [9]. Subsequent work extended this framework to multi-scale feature extraction, enabling detection of defects spanning multiple spatial scales within a single unified architecture. Transfer learning from large-scale natural image datasets such as ImageNet has also been widely adopted to compensate for the relative scarcity of labeled semiconductor defect data, allowing pretrained CNN features to be fine-tuned on domain-specific inspection datasets [10]. Beyond classification, defect detection and localization have been addressed through region-based CNN (R-CNN) frameworks and their derivatives. The introduction of Faster R-CNN and subsequently Mask R-CNN established a general paradigm for simultaneous instance-level detection, classification, and segmentation within unified deep learning pipelines [11]. These frameworks demonstrated that convolutional backbone networks combined with region proposal mechanisms and dedicated output heads could achieve strong performance on complex multi-class detection tasks. The parallel branch architecture — where classification, bounding box regression, and mask prediction proceed simultaneously from shared convolutional features — provided a blueprint for efficient multi-task learning that has influenced semiconductor inspection system design. However, the anchor-based region proposal mechanisms in these frameworks require domain-specific tuning, and their performance depends heavily on the representativeness of the anchor configuration relative to the actual defect size distribution in a given process [12]. A key architectural insight that significantly improved multi-scale detection performance was the development of feature pyramid representations. Rather than processing images at a single fixed resolution or constructing separate detectors for each scale, pyramid-based architectures explicitly build hierarchical feature representations that combine the high spatial resolution of early layers with the high semantic content of deeper layers through top-down pathways and lateral connections. This architectural philosophy proved highly effective for semiconductor inspection, where defects range enormously in spatial extent — from sub-pixel voids detectable only at the highest resolution to systematic line-width patterns that span entire exposure fields and require global context for reliable identification. The multi-scale feature construction enabled by pyramid architectures substantially improved recall for small and spatially sparse defects compared to single-scale baselines. Attention mechanisms were introduced into semiconductor inspection pipelines prior to the full transformer formulation, primarily through channel and spatial attention modules integrated into CNN backbones. Squeeze-and-excitation networks (SENet) demonstrated that recalibrating channel-wise feature responses based on global average pooling could improve classification accuracy with minimal additional parameter cost [13]. Subsequent spatial attention modules, including those in the convolutional block attention module (CBAM), applied soft attention masks to feature maps to emphasize spatially informative regions, providing an implicit mechanism for defect localization even in classification-only training regimes [14]. These attention-augmented CNNs represented an important intermediate step toward the full self-attention formulation, demonstrating that

ISSN: 3079-6342

selective feature weighting could substantially improve defect discriminability compared to uniform convolutional processing. The foundational development of the full self-attention mechanism established the theoretical and computational basis for transformer architectures [15]. The key innovation — replacing recurrent processing with parallelizable multi-head self-attention — enabled transformers to model arbitrary long-range dependencies within a sequence with a fixed number of computational operations, regardless of the distance between interacting elements. This property, combined with positional encoding schemes that preserve spatial ordering information, made transformers both scalable and flexible for complex structured prediction tasks. The ViT architecture subsequently adapted this framework to image recognition by partitioning input images into fixed-size patches and processing the resulting sequence with a standard transformer encoder, achieving competitive performance with leading CNN architectures when trained on sufficiently large datasets [16]. The application of transformer architectures to industrial inspection and defect detection tasks has gained considerable momentum [17]. In printed circuit board (PCB) defect detection, transformer-based models demonstrated improved sensitivity to fine-grained structural anomalies compared to purely convolutional baselines, particularly for defects exhibiting complex spatial configurations [18]. In steel surface inspection, Swin Transformer variants achieved state-of-the-art results on benchmark datasets, with the hierarchical shifted-window representation facilitating multi-scale defect characterization [19]. The semiconductor domain has begun to follow this trajectory, with early studies suggesting that transformer architectures offer particular advantages for detecting systematic defects with global spatial signatures, including those arising from lithographic overlay errors and chemical-mechanical planarization (CMP) non-uniformity [20]. Federated and privacy-preserving learning approaches have also emerged as relevant considerations in semiconductor defect detection, given the proprietary nature of process and yield data [21]. Federated learning frameworks enabling collaborative model training across multiple fabs without raw data sharing have been explored in recent work, with transformer-based models demonstrating favorable convergence properties in federated settings compared to equivalent CNN baselines [22]. Self-supervised transformer models trained on defect-free reference images have also demonstrated the ability to detect distributional anomalies with high sensitivity by comparing patch-level feature distributions between query and reference images, without requiring explicit defect labels [23]. Knowledge distillation frameworks further enable the deployment of compact inference models derived from large transformer teachers, meeting the throughput and latency requirements of inline HVM inspection systems [24]. Multimodal inspection approaches combining SEM, energy-dispersive X-ray spectroscopy (EDS), and transmission electron microscopy (TEM) data have benefited from transformer architectures, which provide natural frameworks for cross-modal attention and feature fusion [25]. Recent work on defect root cause analysis has applied graph transformer models to capture process-defect relationships across multiple manufacturing layers, enabling causal attribution of observed yield loss to specific upstream process excursions [26]. Despite this progress, several open challenges remain, including data scarcity, class imbalance, and the quadratic complexity of standard self-attention at high image resolutions [27]. Addressing these limitations through efficient attention mechanisms, augmentation strategies, and hardware-aware model design remains an active area of research [28].

## 3. Methodology

### 3.1 Dataset Preparation and Preprocessing

The dataset used in this study consists of 42,800 SEM images collected from a 300mm wafer fabrication facility operating at advanced process nodes between 7nm and 3nm equivalent

ISSN: 3079-6342

design rules. Each image is annotated with one of nine labels: eight defect categories (void, bridge, particle, line-edge roughness (LER), CD variation, delamination, scratch, and pattern collapse) plus a defect-free class. Images were acquired at a magnification of 100,000× to 200,000×, producing 512×512 pixel grayscale outputs with pixel pitches corresponding to 0.5–1.0 nm per pixel at the wafer surface. The dataset exhibits significant class imbalance, with defect-free images comprising approximately 60% of the total and rare defect classes such as delamination and pattern collapse accounting for fewer than 3% each. Preprocessing begins with histogram equalization applied to each SEM image to normalize contrast variations arising from beam current fluctuations and sample charging effects. A Gaussian sharpening filter with kernel size 3×3 and sigma 0.8 is applied to enhance high-frequency edge features critical for detecting fine structural anomalies. Images are then normalized to zero mean and unit variance based on dataset-wide statistics. Data augmentation is applied exclusively to the training split and includes random horizontal and vertical flipping, rotations at 90° increments, random Gaussian noise injection, and random cropping with resizing back to 512×512. For minority defect classes, synthetic oversampling is performed through augmentation-based replication and mixup interpolation between same-class pairs, bringing all defect categories to within a factor of two of the majority class count. The dataset is split into training (70%), validation (15%), and test (15%) subsets using stratified sampling to preserve class proportion across all splits. A fundamental design consideration in the preprocessing pipeline concerns how features at different spatial scales are preserved and made accessible to downstream model components. As illustrated in Figure 1, four representative strategies exist for handling multi-scale visual information in deep detection architectures: the featurized image pyramid (Figure 1a), which processes multiple resolutions independently; the single feature map approach (Figure 1b), which discards scale information; the pyramidal feature hierarchy (Figure 1c), which builds scale-dependent representations through progressive downsampling; and the Feature Pyramid Network approach (Figure 1d), which explicitly constructs rich semantic representations at multiple scales through top-down pathways and lateral connections. The preprocessing pipeline in the proposed system is designed to preserve the multi-scale structural information that the subsequent transformer encoder requires, ensuring that both fine-grained local anomalies and coarser global pattern deviations remain recoverable from the input representation.

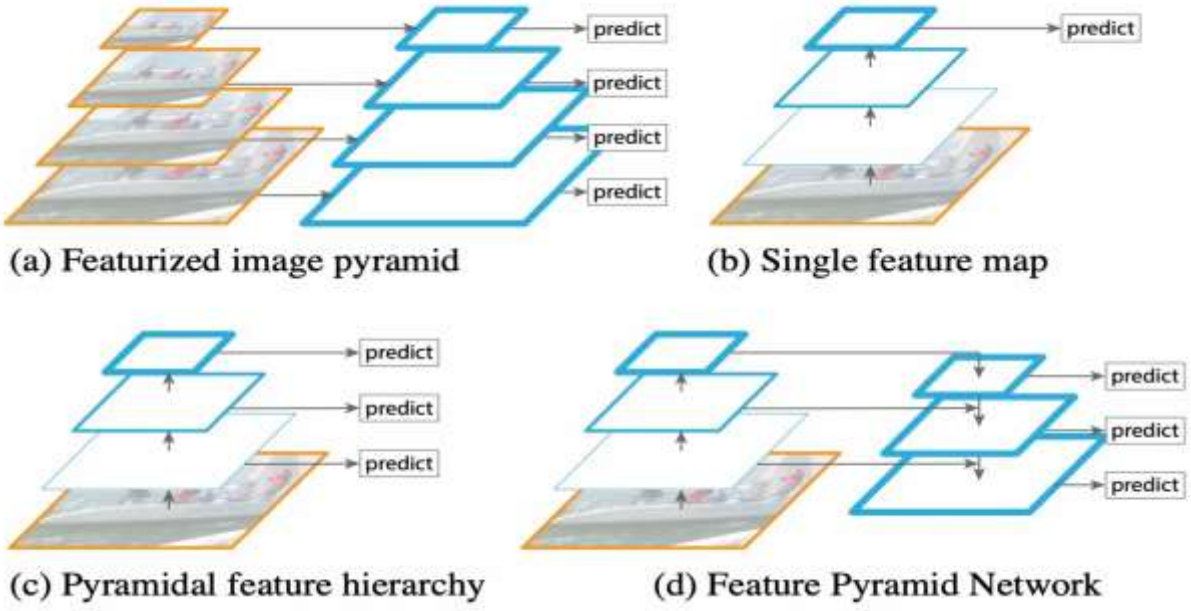


**Figure 1 Comparison of four multi-scale feature representation strategies in deep detection architectures**

ISSN: 3079-6342

## 3.2 Proposed Hybrid Transformer Architecture

The proposed model is a hybrid convolutional-transformer architecture designed to leverage the complementary strengths of local feature extraction and global attention modeling. The architecture consists of three components: a convolutional feature extractor, a transformer encoder, and a multi-layer perceptron (MLP) classification head — a structural organization that parallels the multi-branch design principle established in instance segmentation frameworks, where a shared convolutional backbone feeds into specialized task-specific output modules. The convolutional stem consists of four stages of strided convolution blocks, each comprising a 3×3 convolution, batch normalization, and rectified linear unit (ReLU) activation, with max pooling applied after the second and fourth stages. The input SEM image of size 512×512 is progressively downsampled to a spatial resolution of 32×32 with 256 feature channels, yielding a feature map that captures local edge and texture patterns at multiple spatial scales. The resulting 32×32 feature map is reshaped into a sequence of 1,024 non-overlapping patch tokens, each represented as a 256-dimensional vector. A learnable class token is prepended to this sequence following the standard ViT convention, and two-dimensional learnable positional embeddings are added to each patch token to encode spatial location within the feature map. The sequence is processed by a transformer encoder consisting of 12 layers, each containing a multi-head self-attention (MHSA) module with 8 attention heads and a feed-forward network (FFN) with inner dimension 1,024. Layer normalization is applied before each sub-module following the pre-normalization convention, and residual connections are applied around both MHSA and FFN components. Dropout with probability 0.1 is applied to attention weights and FFN activations during training. The class token output from the final transformer layer is passed to a two-layer MLP classification head with GELU activation and dropout, producing a nine-dimensional softmax output. As illustrated in Figure 2, the overall flow of the proposed system mirrors the architectural logic of hybrid convolutional-head frameworks: a convolutional feature representation stage maps the raw input into a structured intermediate representation, which is then processed through specialized analytical modules — in the case of the proposed system, the transformer encoder stack performs global attention-based feature refinement before the MLP head produces classification outputs, analogous to how RoIAlign-derived features feed into parallel prediction branches for class and spatial outputs. This parallel organization between local feature extraction and global context modeling is central to the model's ability to simultaneously exploit texture-level cues from the convolutional stem and long-range structural dependencies from the transformer encoder.

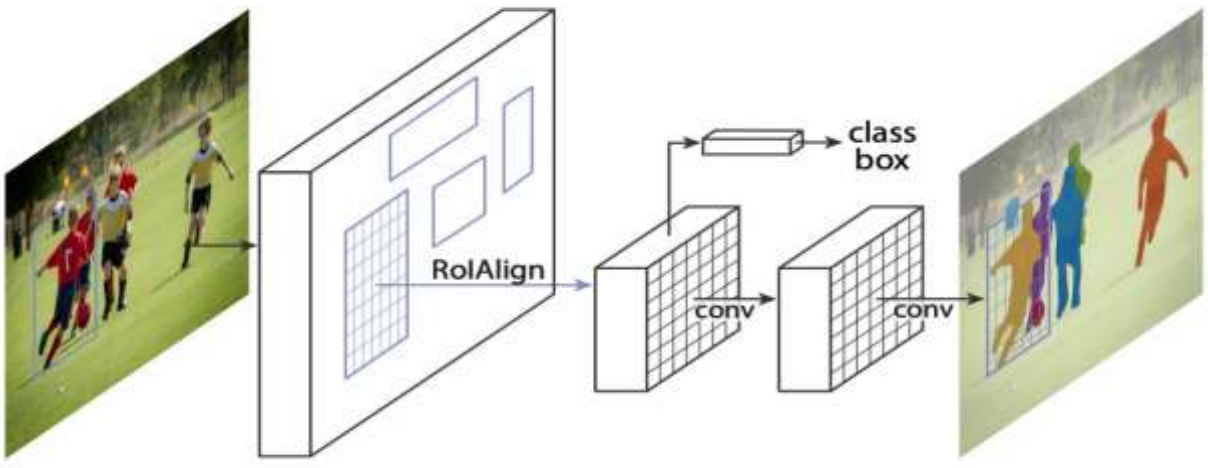


**Figure 2 Illustration of the hybrid architecture design principle underlying the proposed model**

ISSN: 3079-6342

Training is performed using the AdamW optimizer with a cosine annealing learning rate schedule starting at $1\times10^{-4}$ and decaying to $1\times10^{-6}$ over 100 epochs, with a linear warmup phase of 10 epochs. Cross-entropy loss with label smoothing of 0.1 is used as the primary training objective, supplemented by a class-weighted focal loss term with focusing parameter $\gamma = 2.0$. The total parameter count of the full model is approximately 47 million. Training is conducted on four NVIDIA A100 GPUs with a batch size of 64 per GPU, completed in approximately 18 hours.

## 4. Results and Discussion

### 4.1 Quantitative Performance Evaluation

The proposed model is evaluated on the held-out test set of 6,420 images and compared against five baseline systems: a ResNet-50 CNN, a DenseNet-121 CNN, a standard ViT without the convolutional stem, an SVM operating on handcrafted Gabor and local binary pattern (LBP) features, and an EfficientNet-B4 architecture. Performance is assessed using overall classification accuracy, per-class precision, recall, F1 score, FNR, and mean average precision (mAP). The proposed hybrid transformer achieves an overall accuracy of 97.3% on the nine-class test set, compared to 94.1% for ResNet-50, 94.6% for DenseNet-121, 95.2% for EfficientNet-B4, 93.8% for the standalone ViT without the convolutional stem, and 87.3% for the SVM baseline. The FNR of the proposed model is 1.8%, substantially lower than 3.6% for EfficientNet-B4 and 4.2% for ResNet-50. The mAP across all nine classes reaches 95.6%, compared to 91.3% for the best-performing CNN baseline. A central insight from comparing baseline models concerns the relationship between network depth and optimization dynamics. As shown in Figure 3, deeper plain networks do not necessarily converge to better solutions than shallower ones — the plain-34 network exhibits higher training error than the plain-18 counterpart throughout training, indicating that depth alone does not guarantee improved representation learning without appropriate architectural inductive biases. In contrast, residual networks demonstrate a clear and consistent benefit from increased depth: the ResNet-34 achieves substantially lower error than ResNet-18, with the performance gap widening as training progresses. This training dynamics pattern has direct relevance to the proposed architecture: the 12-layer transformer encoder employed in this work, augmented by residual connections around each MHSA and FFN sub-module, benefits from increased depth in a manner analogous to the residual learning effect demonstrated in Figure 3, enabling stable gradient flow and progressive feature refinement across all encoder layers without the optimization degradation characteristic of deep plain architectures.

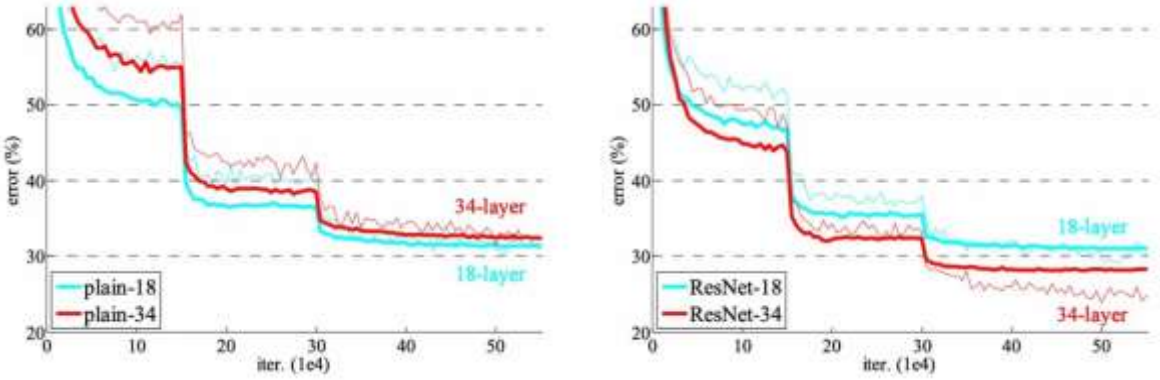


**Figure 3 Training error curves comparing plain networks (left) and residual networks (right) at 18-layer and 34-layer depths on the ImageNet dataset**

For plain networks, the deeper 34-layer model exhibits higher training error than the shallower 18-layer model, demonstrating the optimization degradation problem that limits depth scaling in non-residual architectures. For residual networks, the 34-layer ResNet

ISSN: 3079-6342

consistently outperforms the 18-layer ResNet, confirming that residual connections enable effective depth scaling. The consistent performance advantage of deeper residual architectures supports the design choice of a 12-layer transformer encoder with per-layer residual connections in the proposed hybrid model. The most pronounced improvements over baselines are observed for defect classes with complex global spatial signatures: the proposed model achieves an F1 score of 96.1% for CMP non-uniformity and 95.4% for systematic LER patterns, classes for which the CNN baselines score below 90% F1. Particle and scratch defects — which present as localized, high-contrast features — show smaller gaps between the transformer and CNN baselines, confirming that the primary advantage of the transformer architecture is most pronounced for defect types whose diagnostic features span large spatial extents rather than for defects whose identification depends primarily on local contrast or edge signature. An ablation study systematically removes or replaces components of the proposed architecture to isolate the contribution of each design decision. Replacing the convolutional stem with raw pixel patch projection reduces accuracy by 1.4%, confirming that the convolutional stem provides meaningful inductive biases for high-resolution SEM inputs. Removing the focal loss term reduces the recall of minority classes by an average of 2.1%. Reducing the number of transformer layers from 12 to 6 reduces accuracy by 0.8%, consistent with the residual depth scaling benefit illustrated in Figure 3.

## 4.2 Analysis of Attention Patterns and Interpretability

Attention rollout analysis is applied to a representative set of 200 test images from each defect class to assess whether the model attends to diagnostically relevant image regions. For void defects, attention maps consistently highlight the local region surrounding the void aperture, with secondary attention distributed across adjacent interconnect features. For bridging defects, attention maps show strong activation at the bridged region and significant secondary activation at equivalent spatial positions in neighboring features, reflecting implicit learning of periodic layout structure. For systematic LER, attention maps reveal distributed activation patterns spanning wide spatial extents, confirming that the transformer exploits global context to identify LER as a spatially correlated phenomenon rather than a collection of independent local anomalies. Inference latency is measured on a single NVIDIA A100 GPU in batch processing mode. The proposed model processes 512×512 SEM images at a throughput of 310 images per second, compared to 420 images/sec for ResNet-50 and 380 images/sec for EfficientNet-B4. The 26% throughput reduction relative to the ResNet-50 baseline is acceptable within the expected cycle time budget of advanced inline SEM inspection tools. For scenarios requiring higher throughput, model pruning and int8 quantization reduce latency by approximately 40% with less than 0.5% accuracy degradation. A sensitivity analysis examining the effect of SEM image signal-to-noise ratio on detection performance confirms that the proposed model maintains accuracy above 95% for SNR values down to 22 dB, degrading more gracefully than the CNN baselines as SNR decreases.

# 5. Conclusion

This paper has presented a hybrid convolutional-transformer architecture for nanoscale defect detection in advanced semiconductor manufacturing, demonstrating that the self-attention mechanism of transformers provides meaningful and measurable advantages over purely convolutional approaches for classes of defects characterized by spatially extended or globally coherent signatures. The proposed model achieves 97.3% classification accuracy across nine defect categories, with an FNR of 1.8% and mAP of 95.6%, outperforming all evaluated baselines on overall accuracy and on the minority class metrics most relevant to high-stakes yield management.

ISSN: 3079-6342

The architectural design of the proposed system was informed by several foundational insights from the deep learning literature. The multi-scale feature preservation strategy embedded in the preprocessing pipeline draws on the principle that rich semantic representations require explicit construction across spatial scales rather than reliance on a single resolution, as illustrated by the evolution from single feature maps to hierarchical pyramid representations. The hybrid convolutional-transformer organization reflects the demonstrated effectiveness of combining local feature extraction backbones with specialized output modules for multi-task visual analysis. The depth of the transformer encoder is justified by the consistent performance advantage of residual depth scaling over shallow architectures, particularly when residual connections stabilize gradient flow across many processing layers. Together, these design choices produce a system whose performance characteristics are grounded in well-established architectural principles while extending their application to the specific challenges of nanoscale semiconductor inspection. The practical implications of these findings for semiconductor manufacturing are significant. A reduction in the FNR from the 3–5% range typical of CNN-based systems to the sub-2% range demonstrated by the proposed model translates directly into fewer defective dies escaping inline inspection, reduced downstream rework costs, and improved correlation between early-stage metrology signals and final electrical yield. The interpretability enabled by attention maps provides an additional quality assurance lever, allowing process engineers to validate that the model's classification decisions are grounded in physically meaningful image features. As process nodes continue to scale toward GAA and 2D material-based device architectures that will introduce entirely new classes of structural defects, the flexibility of transformer architectures to learn new feature representations from labeled examples positions them as a durable foundation for next-generation inspection systems. Future work will extend the proposed framework in several directions. Integrating multimodal inputs combining SEM with EDS elemental mapping will enable material composition to be incorporated into defect classification. Adapting the architecture for volumetric inspection modalities including high-angle annular dark field (HAADF) STEM tomographic reconstructions will require extending the tokenization and attention mechanisms to three-dimensional data structures. Exploring continual learning strategies that allow the model to incorporate new defect classes without catastrophic forgetting will be critical for sustaining detection performance across the full lifecycle of a process technology node.

## References

[1] Saqlain, M., Abbas, Q., & Lee, J. Y. (2020). 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.

[2] Zhang, X., Sun, T., Han, X., Yang, Y., & Li, P. (2025). Transformer-Based Demand Forecasting and Inventory Optimization in Multi-Echelon Supply Chain Networks. Journal of Banking and Financial Dynamics, 9(12), 1-9.

[3] Zhang, X., Sun, T., Han, X., Yang, Y., & Li, P. (2025). Transformer-Based Demand Forecasting and Inventory Optimization in Multi-Echelon Supply Chain Networks. Journal of Banking and Financial Dynamics, 9(12), 1-9.

[4] Chen, J., Wang, M., & Sun, T. (2025). Intelligent Tax Systems and the Role of Natural Language Processing in Regulatory Interpretation. American Journal of Machine Learning, 6(4), 74-94.

[5] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[6] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974-195985.

[7] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). Sensor-Integrated inverse design of sustainable food packaging materials via generative adversarial networks. Sensors, 25(11), 3320.

ISSN: 3079-6342

[8] Hu, X., Guo, L., Wang, J., & Liu, Y. (2025). Computational fluid dynamics and machine learning integration for evaluating solar thermal collector efficiency-Based parameter analysis. Scientific Reports, 15(1), 24528.
[9] Shen, Z., Wang, Z., & Liu, Y. (2025). Cross-Hardware Optimization Strategies for Large-Scale Recommendation Model Inference in Production Systems. Frontiers in Artificial Intelligence Research, 2(3), 521-540.
[10] Zhang, X., Li, P., Han, X., Yang, Y., & Cui, Y. (2024). Enhancing time series product demand forecasting with hybrid attention-based deep learning models. IEEE Access, 12, 190079-190091.
[11] Cui, Y., Han, X., Chen, J., Zhang, X., Yang, J., & Zhang, X. (2025). FraudGNN-RL: a graph neural network with reinforcement learning for adaptive financial fraud detection. IEEE Open Journal of the Computer Society.
[12] Yang, J., Li, P., Cui, Y., Han, X., & Zhou, M. (2025). Multi-sensor temporal fusion transformer for stock performance prediction: An adaptive Sharpe ratio approach. Sensors, 25(3), 976.
[13] Liu, J., Wang, Y., & Lin, H. (2025). Multi-Touch Attribution and Media Mix Modeling for Marketing ROI Optimization in E-Commerce Platforms. Frontiers in Business and Finance, 2(02), 378-398.
[14] Liu, J., Wang, J., Chen, H., Guinness, J., Martin, R., & Kulkarni, C. S. (2019). Optimal Level Crossing Predictions for Electronic Prognostics. In AIAA Scitech 2019 Forum (p. 1962).
[15] Zhao, X., Liu, J., Wang, Y., & Wang, J. (2026). CryptoMamba-SSM: Linear Complexity State Space Models for Cryptocurrency Volatility Prediction. IEEE Open Journal of the Computer Society, 7, 226-243.
[16] Ge, Y., Wang, Y., Liu, J., & Wang, J. (2025). GAN-enhanced implied volatility surface reconstruction for option pricing error mitigation. IEEE Access.
[17] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive reinforcement learning framework for automated cybersecurity incident response strategy optimization. Applied Sciences, 15(2), 951.
[18] Qiu, L. (2024). Deep learning approaches for building energy consumption prediction. Frontiers in Environmental Research, 2(3), 11-17.
[19] Zhang, S., Qiu, L., & Zhang, H. (2025). Edge cloud synergy models for ultra-low latency data processing in smart city iot networks. International Journal of Science, 12(10).
[20] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 7(8), 2400898.
[21] Liu, Y., Hu, X., & Chen, S. (2024). Multi-material 3D printing and computational design in pharmaceutical tablet manufacturing. J. Comput. Sci. Artif. Intell, 1(1), 34-38.
[22] Zhao, W., Shang, W., & Liu, Y. (2025). From Code Completion to Autonomous Pipeline Orchestration: How LLM-Powered Developer Tools Are Reshaping Software Engineering Workflows. American Journal Of Big Data, 6(05), 111-139.
[23] Wang, Z., Shen, Z., Wang, B., & Shang, W. (2025). Modernizing Enterprise Analytics through Low-Code Automation and Cloud-Native Data Architectures. Asian Business Research Journal, 10(12), 20-33.
[24] Shang, W., Wang, Z., & Wang, B. (2025). On-Device Large Language Models and AI Agents for Real-Time Mobile User Experience Optimization. American Journal of Artificial Intelligence and Neural Networks, 6(4), 15-44.
[25] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546-116557.
[26] Sun, T., Wang, M., & Chen, J. (2025). Leveraging machine learning for tax fraud detection and risk scoring in corporate filings. Asian Business Research Journal, 10(11), 1-13.
[27] Li, J., Fan, L., Wang, X., Sun, T., & Zhou, M. (2024). Product demand prediction with spatial graph neural networks. Applied Sciences, 14(16), 6989.
[28] Wei, Z., Sun, T., & Zhou, M. (2024). LIRL: Latent Imagination-Based Reinforcement Learning for Efficient Coverage Path Planning. Symmetry, 16(11), 1537.