JM3 Journal of Micro/Nanopatterning, Materials, and Metrology

RESEARCH PAPER

Deep-learning-driven automation of inner and outer object segmentation in SEM/TEM imaging for semiconductor metrology

Isaac Wilfried Sanou,[a,b,*] Julien Baderot,[b] Vincent Barra,[a] Ali Hallal,[b]
Léo Mazauric,[a,b] and Johann Foucher[b]
[a]Université Clermont-Auvergne, CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, Clermont-Ferrand, France
[b]Pollen Metrology, Moirans, France

ABSTRACT. **Background**: In the semiconductor industry, accurate and fast segmentation and metrology on scanning electron microscopy (SEM) and transmission electron microscopy (TEM) images are critical for analyzing complex structures. Traditional image analysis often struggles with low contrast, weak or ambiguous boundaries, and the need to distinguish multiple materials within an object.

Aim: To address these challenges, we present a deep-learning approach that improves segmentation accuracy and automates metrology for inner and outer objects at advanced manufacturing steps.

Approach: We fine-tune the OneFormer architecture with domain-specific adaptations: (i) a combined loss (Boundary, Dice, Binary Cross-Entropy) to sharpen edges and balance classes; (ii) a two-stage training schedule (semantic then panoptic) to stabilize inner/outer delineation; (iii) modality-specific post-processing—shadow handling for SEM and edge enhancement for TEM; and (iv) training on a curated SEM/TEM dataset with expert annotations.

Results: On held-out data, our method reaches 89% recall and 80% precision for object detection (baseline from our previous work: 10% recall, 8% precision). For metrology, we obtain dimensionless R^2 scores of 0.84 (inner height), 0.92 (outer height), 0.82 (inner width), and 0.90 (outer width), outperforming OneFormer without fine-tuning ($R^2 \leq 0.47$) and prior methods ($R^2 \leq 0.57$). The pipeline requires only a global bounding box initialization, whereas classical snakes need multi-step initialization.

Conclusions: These results indicate reliable automated measurement across SEM and TEM, with explicit handling of shadows and weak boundaries. The framework supports semi-automatic operation with user feedback and can be deployed fully automatically after validation, offering a scalable solution for semiconductor metrology.

[DOI: 10.1117/1.JMM.24.4.044201]

Keywords: inner and outer segmentation; metrology; scanning electron microscopy/transmission electron microscopy; annotation; deep learning; Vision Transformer

Paper 25029SS received Mar. 21, 2025; revised Sep. 8, 2025; accepted Sep. 16, 2025; published Oct. 6, 2025.

*Address all correspondence to Isaac Wilfried Sanou, isaac.sanou@pollen-metrology.com, isaacwilfried@gmail.com

2708-8340/2025/$28.00 © 2025 SPIE

1 Introduction

In the field of research and development, engineers strive to enhance industrial processes to produce objects with precise dimensions, high uniformity, and optimal speed and efficiency. The integration of deep learning models has significantly transformed and accelerated these tasks.[1–4] These models provide a unique combination of accuracy, robustness, and speed when appropriately trained with relevant data. However, the effective deployment of deep learning models comes with challenges, particularly in the acquisition of high-quality training data.

The process of acquiring annotated data requires the expertise of skilled users who can accurately delineate the boundaries of objects in a large number of images. This step is critical, as the quality of annotations directly affects the model's ability to learn and generalize. Once the data are acquired and the models are trained, their application in metrology becomes essential. Segmentation, a fundamental task in image analysis, involves partitioning an image into distinct regions to identify and measure specific structural features. In the context of metrology, segmentation allows the extraction of precise measurements from segmented regions, which is crucial for research and development.

Microscopy images generated through scanning electron microscopy (SEM) and transmission electron microscopy (TEM) pose unique challenges. These images often suffer from low contrast boundaries, shadow artifacts, and pronounced textures, as shown in Fig. 1. In addition, segmenting inner objects, such as voids or material inclusions within semiconductor structures, introduces further complexity. These inner objects are often surrounded by materials with similar intensity levels, making boundary detection particularly difficult. Moreover, the intricate shapes and small dimensions of these objects demand a high degree of precision and robustness from segmentation models.

Addressing these challenges requires clear definitions of the core concepts involved. In the context of this study, segmentation refers to the process of partitioning an image into meaningful regions. Outer segmentation focuses on detecting the external structures, such as the primary boundaries of semiconductor components, whereas inner segmentation targets internal elements such as voids or inclusions embedded within these structures. Microscopy, specifically SEM and TEM, is essential for semiconductor imaging. SEM generates high-resolution surface images using an electron beam, whereas TEM offers atomic-scale resolution by transmitting electrons through a sample, enabling the observation of internal structures.[5] Metrology in the semiconductor industry is the set of techniques used to measure dimensions, shapes, and other structural characteristics of fabricated components, ensuring compliance with design specifications.[1] Finally, deep learning refers to a subset of artificial intelligence that uses multilayered neural networks to automatically extract and learn complex features from data, excelling in tasks such as classification, detection, and segmentation.[6]

The adoption of deep learning models for segmentation in the semiconductor domain is crucial to addressing these challenges. These models not only improve measurement precision but also automate processes that traditionally required extensive manual intervention.

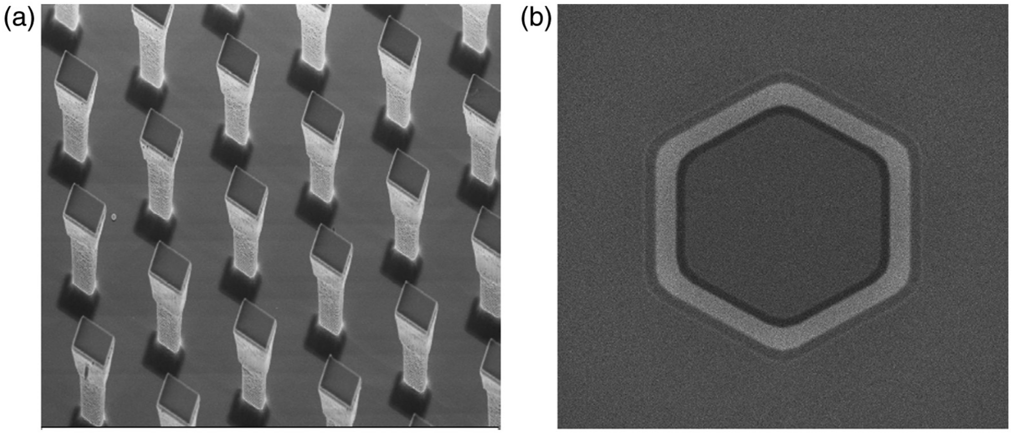

Fig. 1 Examples of the low contrast images. (a) Microwire with some artifacts such as shadows. (b) Simulation of a 3D memory SEM image with low contrast boundaries.

Furthermore, their ability to generalize to novel and complex images is a significant advantage in industrial applications.

In this paper, we propose a deep learning model specifically designed to tackle the unique challenges of SEM/TEM image segmentation and metrology. Our approach incorporates advanced loss functions tailored to address problems related to low contrast, complex boundaries, and multi-material compositions. A tool developed in prior work[7] was employed to annotate specific structures, facilitating the creation of high-quality training datasets. This paper extends our preliminary conference paper[8] by incorporating (1) comprehensive ablation studies on loss components, (2) explicit differentiation between SEM and TEM challenges, (3) quantitative comparisons with classical segmentation methods, and (4) extensive validation on a larger dataset.

The remainder of this paper is organized as follows: In Sec. 2, we review the classical approaches based on standard algorithms and recent methods utilizing deep learning. In Sec. 3, we detail our proposed approach for inner and outer segmentation. Section 4 presents the experimental results and discusses the effectiveness of our model for segmentation and metrology. Finally, Sec. 5 concludes the paper and provides directions for future research.

2 State-of-the-Art of Inner and Outer Segmentations

Accurate segmentation of SEM/TEM images has long been a critical task in the semiconductor industry due to its importance in metrology and structural analysis. Segmenting SEM/TEM images is incorporated in this study, not only outer objects but also inner objects, as shown in Fig. 2.

In this section, we first analyze SEM versus TEM imaging challenges. We then review traditional methods for inner and outer segmentations and state-of-the-art deep learning approaches, highlight their limitations, and identify the gaps that our proposed framework seeks to address.

2.1 SEM/TEM-Specific Challenges for Automated Metrology

SEM and TEM imaging present distinct challenges for automated metrology due to their fundamentally different physical principles and resulting image characteristics. Understanding these differences, as shown in Fig. 3, is crucial for developing robust segmentation algorithms. Concerning SEM-specific challenges:

- Shadows from tilted views: Surface topography creates shadows that can be mistaken for material boundaries.[5]
- Charging artifacts: Non-conductive samples accumulate charge, creating bright or dark regions unrelated to material composition.

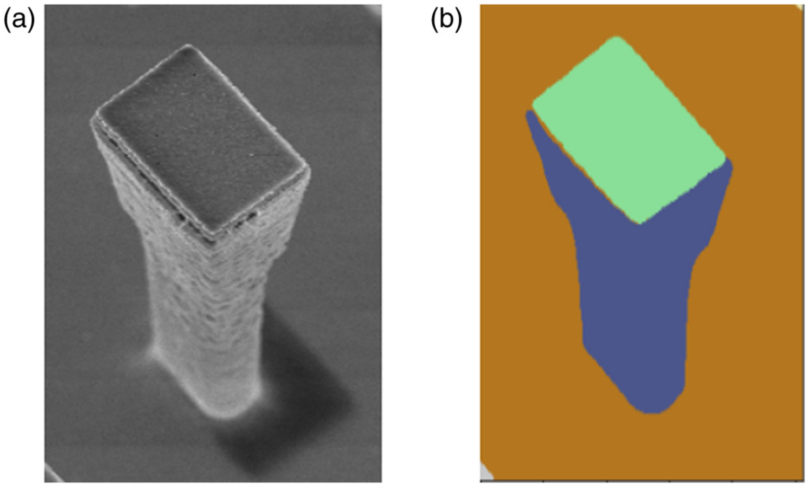

Fig. 2 (a) Original microwire image. (b) Segmentation of outer object (in blue) and inner object (in green).

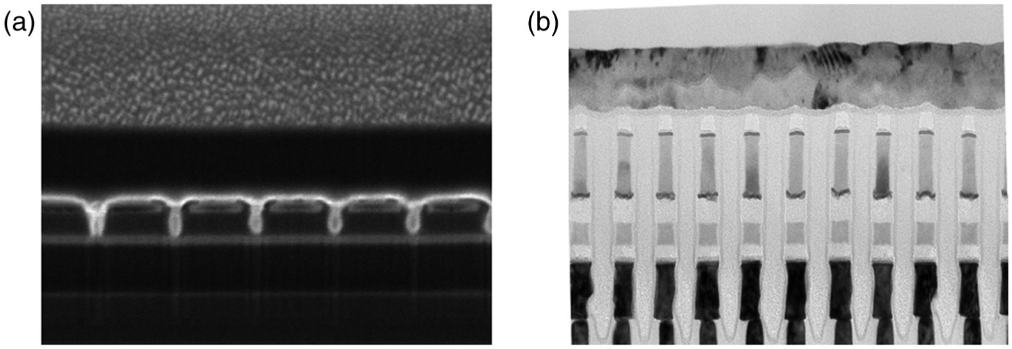

Fig. 3 Comparison of SEM and TEM challenges with different artifacts. (a) SEM image. (b) TEM image.

- Edge effects: Sharp edges appear brighter due to increased secondary electron emission.
- Limited depth of field: Can cause contour discontinuities in three-dimensional (3D) structures.

For TEM-specific challenges, we can notice the following:

- low signal-to-noise ratio: especially for thin specimens or beam-sensitive materials[9]
- mass–thickness contrast: creates ambiguous boundaries among materials of similar atomic number
- beam-induced drift: causes image distortions during acquisition.

2.2 Classical Approaches

Classical segmentation techniques have been extensively studied and remain foundational in image analysis. These methods typically rely on deterministic algorithms and mathematical models to delineate object boundaries based on pixel intensity, gradients, or energy minimization. Below, we outline the most prominent approaches:

2.2.1 *Energy minimization methods*

One of the earliest and most widely used approaches involves energy-based models, such as active contours (snakes).[10] Active contours deform an initial contour toward object boundaries by optimizing an energy function composed of internal and external forces. Variants of these methods, such as snakes combined with local shape models[11] and region-based active contours,[12] incorporate local image information to improve robustness in challenging scenarios. These methods are often sensitive to initialization and noise, leading to sub-optimal results in complex images such as SEM/TEM where boundaries are faint or objects have intricate shapes.

2.2.2 *Watershed algorithms*

The watershed algorithm[13] is a popular gradient-based segmentation technique that treats an image as a topographical surface. It partitions the image into regions by simulating the flooding of the surface from seed points. Improved variants[14] aim to reduce over-segmentation, a common drawback when objects have similar intensities or when noise is present. Although computationally efficient, watershed algorithms often struggle with low-contrast images, as seen in SEM/TEM data, and require preprocessing steps such as gradient smoothing or thresholding to reduce sensitivity to noise.

2.2.3 *Graph cuts*

Graph-based methods, such as graph cuts,[15,16] model segmentation as an optimization problem where pixels or regions are nodes in a graph. The edges represent the similarity among pixels, and the segmentation is obtained by finding the minimum cut that separates the graph into distinct regions. These methods are powerful for global optimization but can be computationally

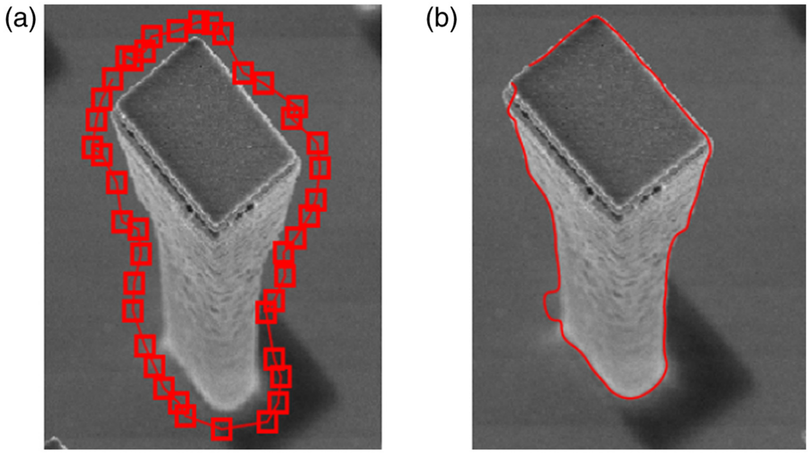

Fig. 4 (a) Original microwire image. (b) Segmentation using the classical method.

intensive for large images. They also require carefully designed cost functions to handle complex textures and low-contrast boundaries.

2.2.4 *Edge detection and thresholding*

Edge detection techniques, such as the Canny edge detector and Otsu's thresholding, are widely used for basic segmentation tasks.[17,18] These methods detect object boundaries based on changes in pixel intensity or gradients. However, in SEM/TEM images, these techniques are prone to over-segmentation and noise sensitivity, especially when the objects and background have similar intensities.

Despite their widespread use, classical methods face several limitations:

- Initialization sensitivity: Poor starting conditions in methods such as active contours can lead to incomplete or incorrect segmentation, as shown in Fig. 4.
- Noise and contrast sensitivity: Low-contrast boundaries and noisy textures in SEM and TEM images degrade the performance of these algorithms.
- Manual intervention: Many classical approaches require extensive manual tuning or preprocessing, making them unsuitable for high-throughput applications in industrial contexts.

These limitations have motivated the exploration of data-driven approaches, such as deep learning, which can learn features directly from data to overcome these challenges.

2.3 Deep Learning Approaches

Recent advancements in image segmentation have been driven by the rise of deep learning, particularly convolutional neural networks (CNNs) and transformer-based architectures. These methods leverage the power of large-scale data and complex network architectures to achieve state-of-the-art performance in diverse segmentation tasks.

2.3.1 *CNN-based models*

Models such as mask region-based convolutional neural network (R-CNN),[19] polygon RNN++,[20] and deep snake[21] have introduced sophisticated techniques for instance segmentation and boundary refinement. For example, mask R-CNN extends the faster R-CNN framework by adding a parallel branch for pixel-wise object segmentation, enabling the detection and segmentation of individual instances in an image. Deep snake combines boundary prediction with contour deformation, making it particularly effective for objects with intricate shapes. Although these methods demonstrate high performance, their reliance on large annotated datasets can be a limitation in domains such as SEM/TEM imaging, where annotated data are limited.

2.3.2 *Transformer-based models*

Transformer-based architectures, such as Mask2Former,[22] have pushed the boundaries of segmentation by incorporating self-attention mechanisms that capture long-range dependencies.

This allows for more accurate segmentation of complex objects in noisy or low-contrast environments. OneFormer[23] is particularly noteworthy for its ability to unify semantic, instance, and panoptic segmentation tasks within a single framework. This versatility makes it an attractive choice for applications such as SEM/TEM segmentation, where objects can vary widely in size, shape, and contrast.

2.3.3 *Generalized segmentation models*

Meta's Segment Anything Model[24] introduces a generalized framework capable of handling various image types with minimal task-specific tuning. FastSAM,[25] a lightweight derivative, uses YOLOv8[26] for faster segmentation while maintaining reasonable accuracy. These models are designed for broad generalization but may lack the specificity required for tasks involving complex textures and intricate boundaries, as seen in SEM/TEM data.

2.3.4 *Unsupervised and self-supervised methods*

Methods such as Cut and Learn,[27] DINO,[28] and LOST[29] aim to reduce the dependency on annotated data by learning meaningful representations in an unsupervised manner. Although promising, these approaches typically underperform compared with supervised methods, particularly in highly specialized domains such as semiconductor imaging.

Advantages of recent approaches are multiples:

- Robustness: CNNs and transformers are less sensitive to noise and initialization compared with classical methods.
- Feature learning: Deep learning models automatically learn relevant features from data, making them adaptable to diverse tasks and conditions.
- Scalability: Once trained, these models can process large volumes of data quickly, making them ideal for industrial applications.

Despite their advantages, recent approaches come with challenges, such as the need for large annotated datasets, high computational requirements, and domain adaptation for specialized applications. In this work, we adopt OneFormer as the base architecture due to its versatility and strong performance across segmentation tasks. We enhance it further with domain-specific adaptations, including advanced loss functions and fine-tuning strategies, to address the unique challenges of SEM/TEM image segmentation.

3 Proposed Deep Model for Inner and Outer Segmentation

Our proposed deep learning model builds upon the strengths of state-of-the-art transformer-based architectures and is specifically tailored to address the unique challenges of SEM/TEM image segmentation. These challenges include low contrast boundaries, intricate object shapes, and the coexistence of inner and outer objects within complex semiconductor structures. To address these specific challenges, we implement several strategies:

- Comparison: We compare classical approaches for inner and outer objects versus recent approaches based on deep learning.
- Multi-task training: We fine-tune a model from different segmentation modes for a better understanding of inner and outer objects.
- Loss function adaptation: We weight Dice loss (DL), binary cross-entropy loss, and boundary loss (BL) differently to preserve faint internal boundaries while emphasizing outer boundaries annotation in SEM to handle shadow artifacts.
- Enhanced edge processing for TEM: We apply stronger smoothing and leverage multiscale features to better handle weak boundaries during inference.
- Post-processing: We perform global contour refinement adapted to each imaging mode using mathematical morphology.

This section describes the key components of our approach, including the overall architecture, training strategies, loss functions, and adaptations made for the SEM/TEM domain.

3.1 Overall Architecture

Our core model builds on OneFormer, a transformer framework that unifies semantic, instance, and panoptic segmentation within a single architecture. Compared with Mask2Former, a strong transformer-based alternative for example, OneFormer offers three advantages for our setting: (i) unified task handling that aligns with outer and inner delineation, which mixes instance-like and semantic cases; (ii) stronger transfer in our preliminary runs when fine-tuned from generic pretraining (e.g., COCO/ADE20K), despite the domain gap with SEM/TEM; and (iii) the DiNAT backbone, which provides rich multi-scale context and clear boundaries at a moderate memory cost. In a controlled side-by-side comparison (identical schedules, augmentations, and post-processing), OneFormer produced cleaner inner boundaries and fewer shadow-induced errors on representative SEM/TEM subsets.

We extend this architecture to better suit the requirements of SEM/TEM images by incorporating the following enhancements:

- Multi-task training: The model is trained sequentially on semantic segmentation (focusing on outer object detection) and panoptic segmentation (integrating inner object detection). This multi-task strategy enables the model to learn both global and local features, ensuring accurate segmentation of objects with varying sizes and contrasts.
- Feature extraction: The transformer-based encoder leverages self-attention mechanisms to capture long-range dependencies in the image, making it well-suited for handling intricate object boundaries and low-contrast regions commonly found in SEM/TEM data.
- Fine-tuning with SEM/TEM data: The pre-trained OneFormer model is fine-tuned on a domain-specific dataset of SEM/TEM images. Fine-tuning allows the model to adapt its weights to the unique textures and features of microscopy images while reducing the need for extensive labeled data.[30]

The overall pipeline is illustrated in Fig. 5. The architecture's modular design allows for seamless integration of custom loss functions, as discussed below.

3.2 Training Strategy and Data Augmentation

To ensure robust performance under data-constrained conditions, we adopt a training strategy grounded in frugal learning principles.[30] This involves the following:

- Data augmentation: Techniques such as random rotations, flips, contrast adjustments, and Gaussian noise are applied to increase the diversity of the training dataset[31] and improve the model's generalization capabilities.
- Sequential training: The training process begins with semantic segmentation tasks to help the model learn fundamental features, followed by panoptic segmentation tasks that address both inner and outer object detection. This sequential approach enables the model to handle complex structures effectively.

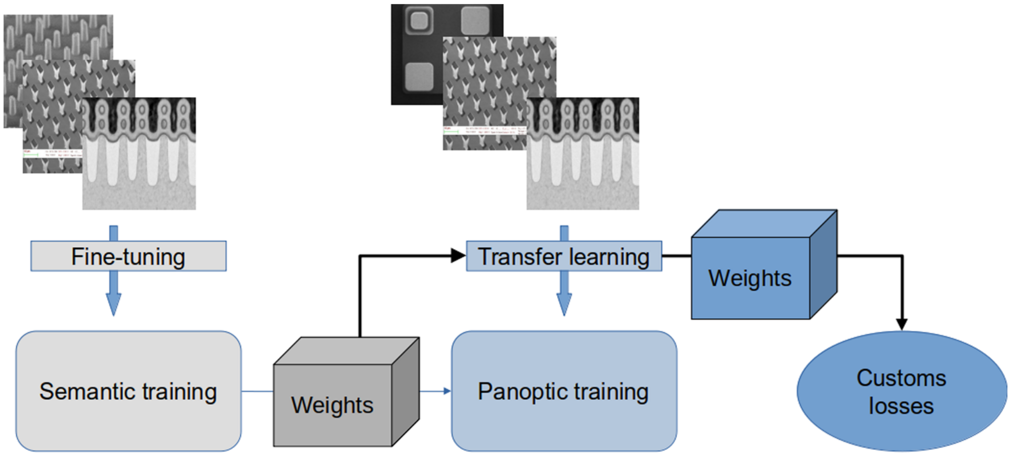

Fig. 5 Our pipeline for inner and outer segmentation.

- Selective weighting: Specific loss weights are assigned to inner and outer objects during training to account for their varying prevalence and importance in SEM/TEM datasets.

3.3 Loss Functions for Enhanced Segmentation

To address the unique challenges of SEM/TEM images, we integrate a suite of advanced loss functions tailored to improve segmentation accuracy and robustness. Each loss function complements the others, targeting specific aspects of the segmentation task:

- BL: This loss function focuses on improving edge detection and ensuring precise boundary delineation. It minimizes discrepancies between the predicted and ground truth boundaries, which is critical for accurately segmenting intricate shapes and fine details. The boundary loss is defined as

$$\mathcal{L}_{\text{BL}} = 1 - \frac{\sum_{i=1}^{N} g_i \cdot p_i}{\sum_{i=1}^{N} g_i + \sum_{i=1}^{N} p_i - \sum_{i=1}^{N} g_i \cdot p_i}, \tag{1}$$

  where g_i and p_i are the ground truth and predicted boundary pixels, respectively, and N is the total number of pixels.
- DL: This loss optimizes the overlap between predicted and ground truth masks, making it particularly effective for segmentation tasks with imbalanced datasets. It is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i}, \tag{2}$$

  where y_i and $\hat{y}_i$ are the ground truth and predicted values for each pixel $\in$ [0,1], respectively.
- BCE loss: This loss function balances class importance, ensuring that the model does not overfit to dominant classes (e.g., background). It is defined as

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{M}\sum_{k=1}^{M}[t_k \ln(p_k) + (1 - t_k)\ln(1 - p_k)], \tag{3}$$

  where t_k denotes the target label (ground truth) for the k'th sample, and p_k corresponds to the predicted probability generated by the model.

3.4 Combined Loss Function

We combine BL, DL, and BCE into a single unified loss function for all segmented objects

$$\mathcal{L}_{\text{total}} = \lambda_{\text{DL}} \cdot \mathcal{L}_{\text{DL}} + \lambda_{\text{BL}} \cdot \mathcal{L}_{\text{BL}} + \lambda_{\text{BCE}} \cdot \mathcal{L}_{\text{BCE}}, \tag{4}$$

where $\lambda_{\text{DL}} = 0.5$, $\lambda_{\text{BL}} = 0.3$, and $\lambda_{\text{BCE}} = 0.2$.

Unified approach for inner and outer objects:

- Importantly, we do not apply different loss weights for inner versus outer objects. Instead:
- All annotated objects (both inner and outer) are treated equally by the loss function.
- The model learns to distinguish and segment both types through the training data, not through loss weighting.
- Images are annotated with both inner objects (voids and inclusions) and outer objects (material boundaries) when present.
- The combined loss naturally handles the varying difficulties of inner versus outer segmentation through its three complementary components.

This unified approach simplifies training while allowing the model to adapt to the specific challenges of each object type through the data itself rather than manual hyperparameter tuning.

3.5 Adaptations for SEM/TEM Image Analysis

The unique characteristics of SEM/TEM images, such as low contrast, intricate textures, and the coexistence of inner and outer objects, require specialized adaptations to ensure accurate segmentation and robust feature extraction. To address these challenges, the proposed model integrates several domain-specific enhancements tailored to the requirements of semiconductor image analysis. Transfer learning plays a crucial role in adapting the proposed model to SEM/TEM images. A pre-trained OneFormer model is fine-tuned on a domain-specific dataset, enabling the model to learn features that are unique to microscopy images. The fine-tuning process involves optimizing the model on a smaller, specialized dataset that captures the variability of semiconductor structures, including differences in material composition, shape, and texture. This approach significantly reduces the need for extensive labeled data while ensuring that the model generalizes effectively to unseen SEM/TEM images. Furthermore, the fine-tuning process incorporates selective weighting of loss functions to prioritize the accurate segmentation of inner objects, which are often more challenging to detect than outer boundaries.

To facilitate accurate metrology, the segmentation result is subjected to a contour extraction step as part of the post-processing pipeline. Once the inner and outer objects are segmented, their contours are extracted using methods tailored to SEM/TEM images, such as edge refinement and morphological operations. This step ensures that the boundaries of segmented objects are sharp and precise, allowing for reliable geometric measurements as shown in Fig. 6. This post-processing step is particularly useful for applications requiring high-precision measurements, such as the analysis of void dimensions or the thickness of material layers.

A key challenge in SEM/TEM image analysis is achieving a balance between the segmentation of inner and outer objects, as these regions often differ in size, contrast, and structural complexity. To address this, the model employs a dynamic loss weighting mechanism that assigns higher weights to inner objects during training. This ensures that the model allocates sufficient attention to capturing small, intricate structures without compromising the accuracy of outer object segmentation. The dynamic weighting is adjusted based on the prevalence and difficulty of each object type within the dataset, allowing the model to adapt to varying image conditions. SEM/TEM images often contain artifacts, such as shadows, reflections, or noise, which can interfere with segmentation accuracy. To mitigate these effects, the model integrates feature normalization techniques and noise-reduction filters within the early stages of the encoder. In addition, the model leverages multi-scale feature extraction to capture both global context and fine details, enabling it to distinguish true object boundaries from noise-induced edges.

The combination of these domain-specific adaptations ensures that the proposed model is well-suited to handle the complexities of SEM/TEM image segmentation. Targeted post-processing techniques permit to the model to achieve high precision in segmenting both inner and outer objects. These adaptations, coupled with domain-specific fine-tuning and robust loss functions, enable the model to perform reliably across diverse semiconductor imaging conditions.

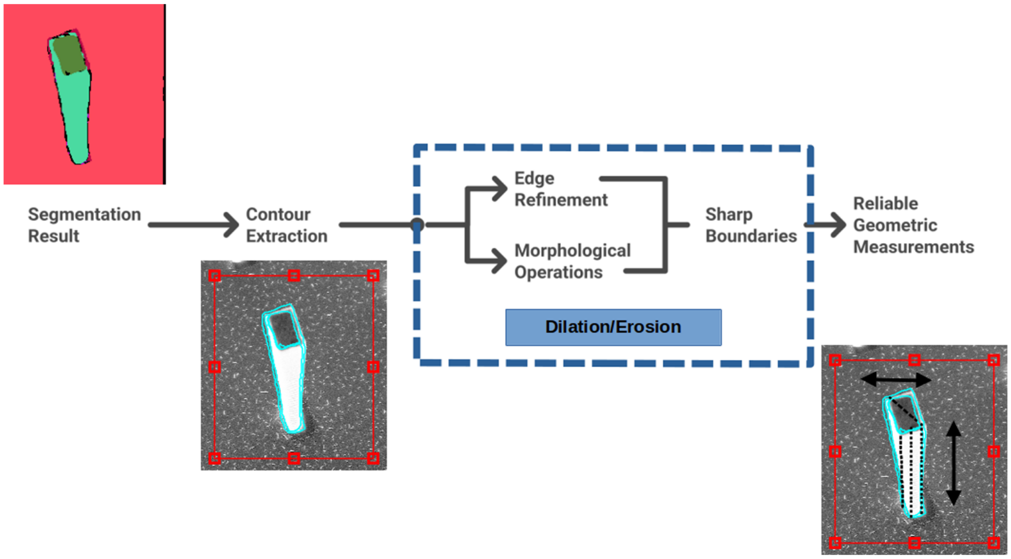

Fig. 6 Our pipeline for contour extraction and metrology.

4 Experiments: Metrology, Results, and Discussion

This section provides details for our dataset and the different metrics to evaluate our approach. We also display results for detecting the inner and outer contours of objects using our proposed post-processing.

4.1 Dataset and Metrics

The proposed deep learning model was trained and evaluated on a specialized dataset of SEM and TEM images, tailored for inner and outer segmentation in semiconductor structures. In the following, we describe the dataset composition, the preprocessing steps, and the evaluation metrics used to assess the model's performance, including the newly introduced precision/recall metrics.

The dataset consists of high-resolution SEM/TEM images annotated to reflect the difficult structures and challenging conditions typically encountered in semiconductor manufacturing. The dataset is composed of the following:

- Training set: 100 annotated images containing a total of 2467 objects. These objects include a mix of inner structures, such as voids and material inclusions, and outer structures, such as boundaries of semiconductor components, as shown in Fig. 7.
- Test set: 15 annotated images containing 421 objects. This test set was designed to cover a wide range of structural variability, including objects with low contrast boundaries, irregular shapes, and varying scales.

The dataset was carefully curated to include both synthetic and real-world images, ensuring sufficient diversity in object shapes, textures, and imaging conditions. Synthetic images were generated to simulate extreme scenarios, such as very thin inner objects or overlapping boundaries, to robustly test the model's generalization capabilities. To standardize the data and enhance model performance, we normalize the pixel intensity values to the range [0, 1] for consistent input to the model.

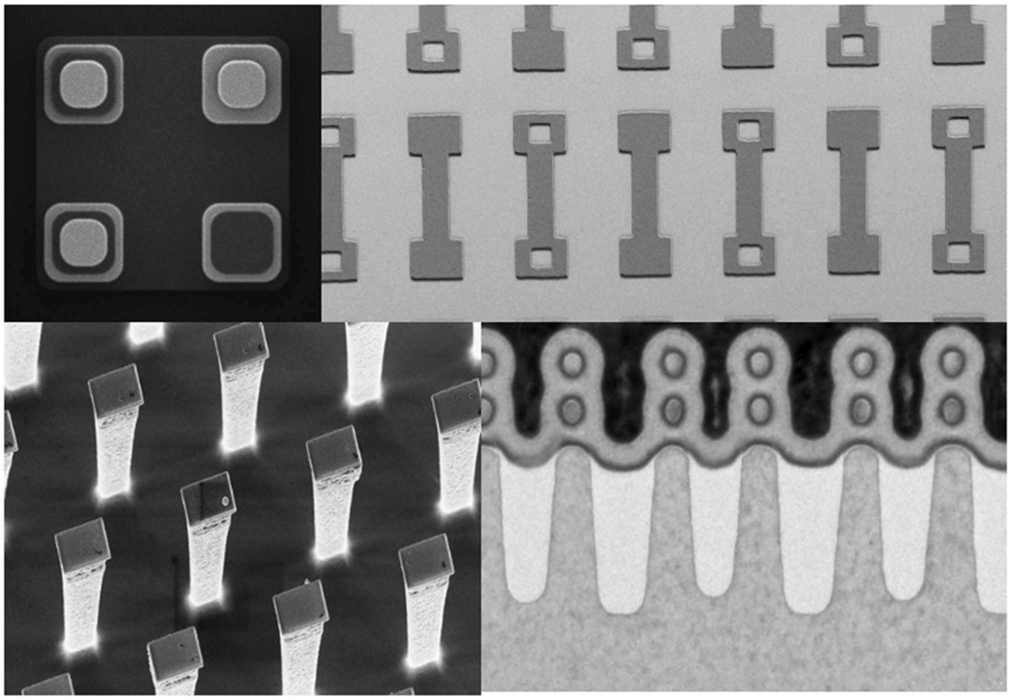

Fig. 7 Samples of the public dataset used. A part of the images comes from the public SEM NFFA-Europe database.[32]

4.1.1 *Evaluation metrics*

To comprehensively evaluate the performance of the model, three key metrics were used: the R^2 score, precision, and recall. Each metric captures different aspects of the segmentation task, as described below.

R^2 score for height and width measurements: The R^2 score evaluates the correlation between the predicted and ground truth measurements of object dimensions, specifically height and width. It is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(m_i^{\text{pred}} - m_i^{\text{true}})^2}{\sum_{i=1}^{n}(m_i^{\text{true}} - \overline{m}^{\text{true}})^2}, \tag{5}$$

where m_i^{pred} and m_i^{true} are the predicted and ground truth measurements (height or width) for object i, respectively, and $\overline{m}^{\text{true}}$ is the mean of the ground truth measurements. This metric highlights the model's accuracy in capturing geometric dimensions, which is crucial for metrology applications. High R^2 scores indicate that the model captures object dimensions accurately, which is crucial for applications requiring precise measurements.

4.1.2 *Precision*

Precision measures the proportion of correctly identified objects relative to the total number of objects predicted by the model. It evaluates the model's ability to avoid false positives, which is critical in ensuring that identified regions correspond to actual objects. Precision is defined as

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}. \tag{6}$$

A high precision value indicates that the model makes fewer incorrect predictions, ensuring that the segmented objects truly belong to the targeted class rather than being artifacts or misclassified structures.

4.1.3 *Recall*

Recall measures the model's ability to correctly identify all relevant objects (inner and outer) in the image. It is particularly important in scenarios where missing objects (false negatives) can have significant implications, such as overlooking voids or material defects. Recall is defined as

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}. \tag{7}$$

A high recall indicates that the model effectively detects a large proportion of the objects present in the dataset, even under challenging imaging conditions. In applications requiring high confidence in segmentation, such as automated metrology, precision ensures that predictions are reliable and consistent.

This comprehensive evaluation ensures that the model not only performs well on individual metrics but also balances the trade-off among the generalization of the task for objects that were not present in the dataset, particularly for challenging SEM/TEM image segmentation tasks.

4.2 Metrology, Results, and Discussion

Our proposed segmentation framework's performance was evaluated by comparing it to two reference models: (1) OneFormer without fine-tuning, which is used as a standard to estimate how well domain-specific adaptation works, and (2) the baseline approach,[7] which is a previously created technique that has been optimized for SEM/TEM image segmentation. The evaluation metrics include precision and recall (% of successfully detected items), as well as R^2 scores for both inner and outer object height and width measurements. The results are summarized in detail in Table 1.

- Height measurements: A low R^2 score of 0.11 was obtained by OneFormer without fine-tuning for inner objects, indicating its limited ability to generalize to SEM/TEM pictures without domain adaptation. This score was raised to 0.41 by the baseline method and

Table 1 Performance comparison among OneFormer without fine-tuning, the baseline approach, and our proposed method.

Metric	OneFormer	Baseline[7]	Our approach
R^2 score for height (inner)	0.11	0.41	0.84
R^2 score for height (outer)	0.46	0.52	0.92
R^2 score for width (inner)	0.13	0.33	0.82
R^2 score for width (outer)	0.47	0.57	0.90
Precision	5%	8%	80%
Recall	6%	10%	89%

R^2 score $\in$ [0; 1]. Precision and recall are expressed as percentages.

considerably raised to 0.84 by our model, demonstrating its capacity to precisely represent the dimensions of interior objects. OneFormer achieved a moderate R^2 score of 0.46 for exterior objects, improving somewhat to 0.52 with the baseline and achieving 0.92 with our method, indicating a significant improvement in boundary delineation.

- Width measurements: Similarly, the R^2 score for inner object width was particularly low for OneFormer without fine-tuning (0.13), showing its difficulty in segmenting narrow structures. The baseline improved the R^2 score result to 0.33, whereas our approach significantly outperformed both, achieving a R^2 score of 0.82. For outer object width, OneFormer attained 0.47, the baseline R^2 score reached 0.57, and our model improved the R^2 score further to 0.90, ensuring robust segmentation across different object scales.
- Object detection accuracy (recall): The number of correctly detected objects increased significantly across models. When applied to OneFormer without fine-tuning, it only detected 6% of objects correctly, demonstrating that pre-trained models struggle when applied directly to SEM/TEM images. The baseline approach improved this to 10%, but our method substantially outperformed both, detecting 89% of objects correctly. This improvement highlights the effectiveness of the proposed model in handling complex structures and varying contrast levels.
- Precision: OneFormer without fine-tuning achieved a low precision of 5%, indicating that many detected objects were false positives. The baseline approach slightly improved this to 8%, but our model exhibited a significant increase to 80%, demonstrating its ability to correctly classify segmented objects and minimize incorrect detection.

These results illustrate the necessity of fine-tuning for domain adaptation. The pre-trained OneFormer model, despite its strong generalization capabilities in world natural image segmentation, fails to effectively segment SEM/TEM images without adaptation. The baseline method, while showing improvement, remains suboptimal compared with our approach, which integrates domain-specific fine-tuning and advanced loss functions to achieve superior segmentation accuracy.

By leveraging advanced loss functions, fine-tuning strategies, and domain-specific adaptations, our model demonstrates substantial improvements in both segmentation accuracy and

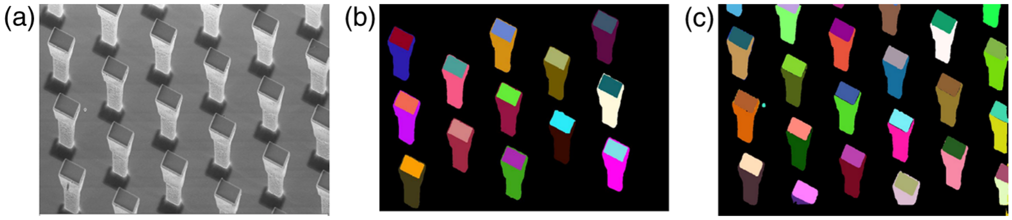

Fig. 8 Example of inference. (a) Original SEM microwire image. (b) Ground truth manually labeled. (c) Inference of our approach.

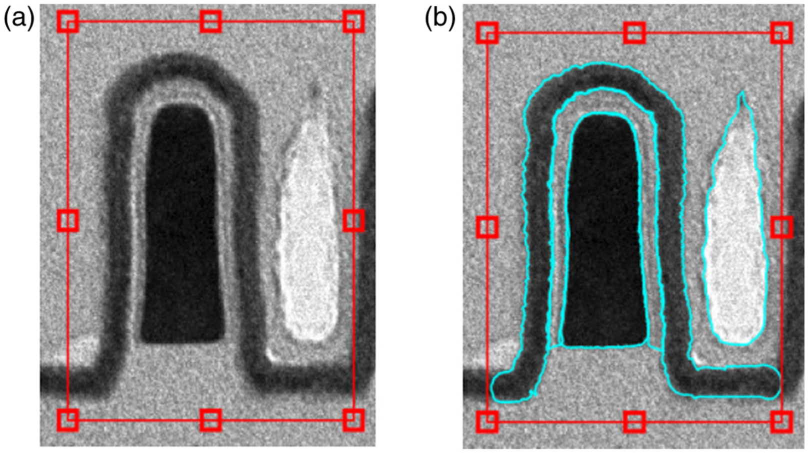

Fig. 9 Example of inference of the object with a bounding box. (a) Original TEM image. (b) Extraction of contours after the inference of our approach with post-processing.

metrology precision. As illustrated in Fig. 8, our approach not only aligns closely with the ground truth but also identifies additional inner and outer objects with high precision. The extracted contours, as shown in Fig. 9, confirm the model's ability to generate well-defined segmentations, which is crucial for metrology applications.

A more detailed breakdown of the performance improvements is presented in Table 1, reinforcing the impact of fine-tuning and domain-specific enhancements.

As shown in Fig. 10, the snake algorithm fails when the image contrast is weak or poorly defined. In this example, it is unable to delineate the inner-object boundary due to low contrast. By contrast, our approach segments both outer and inner objects in this difficult case. It is also

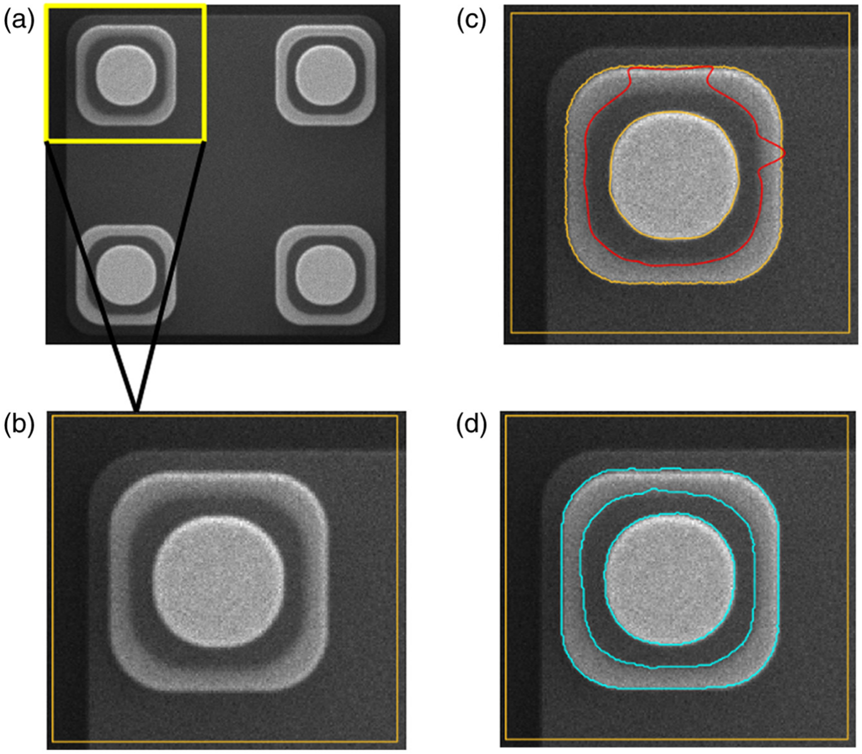

Fig. 10 Comparison between our approach and the snake algorithm on 3D simulated memory. (a) Original image. (b) Cropping of an object with a bounding box. (c) Snake result. (c) Proposed approach result.

Table 2 Ablation study on loss components.

Loss configuration	Precision (%)	Recall (%)	R^2 height	R^2 width
BCE	70	71	0.70	0.72
Dice	75	83	0.76	0.70
BL	77	82	0.81	0.79
BCE + Dice + BL	**80**	**89**	**0.88**	**0.86**

Note: bold indicates the best result in each column.
R^2 scores $\in$ [0,1]. Precision and recall are expressed as percentages.
Combined loss uses weights $\lambda_{\mathrm{BCE}} = 0.2$, $\lambda_{\mathrm{DL}} = 0.5$, and $\lambda_{\mathrm{BL}} = 0.3$.
BL, boundary loss; DL, Dice loss; BCE, binary cross-entropy.

more efficient: it requires only a global bounding box, whereas snake typically needs multi-step initialization.

4.3 Ablation on Loss Components

We quantify the contribution of each loss term to SEM and TEM subsets. Table 2 reports precision, recall, and metrology R^2 (height/width). Compared with BCE alone, Dice improves recall on small inner structures, whereas BL brings the largest gains in metrology fidelity (higher R^2). Unless stated otherwise, all runs share the same schedule, augmentations, and post-processing.

4.4 Limits and Discussion

Despite the strong performance of our model, some challenges remain, particularly in the presence of background granules, which can be misclassified as separate objects. Figure 11 illustrates inference outputs for both semantic and instance segmentations on an image containing different materials. In the instance segmentation case, the model exhibits sensitivity to background

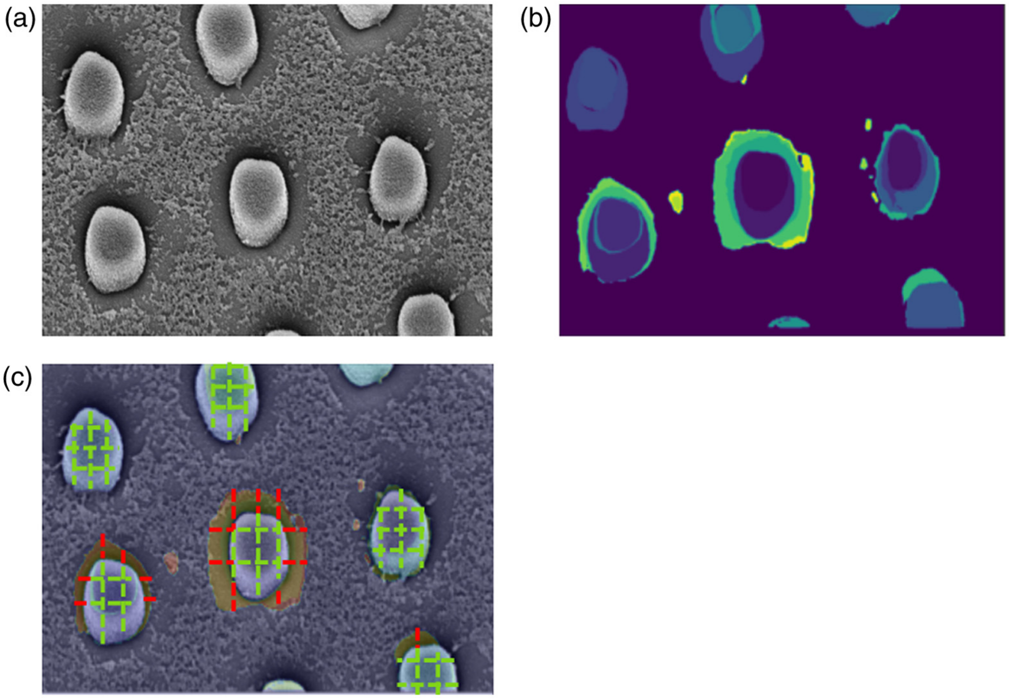

Fig. 11 Example inference on a SEM image containing multiple materials. (a) Original image. (b) Instance segmentation. (c) Overlay of the segmentation on the original image. The shadow is delineated by a red dash–dot line and is excluded from metrology; corrected measurements on the material boundary are indicated by a green dotted line. Detecting and separating shadows is reported here as a current limitation and a direction for future work.

granules, potentially leading to over-segmentation. However, it remains capable of accurately identifying internal structures of interest. This issue can be mitigated through post-processing steps, such as morphological filtering, or by incorporating a background suppression mechanism within the attention layers.

To further improve segmentation precision, future work will focus on integrating advanced attention mechanisms, such as multidimensional collaborative attention[33] and contrast-aware attention,[34] particularly for handling low-contrast or highly textured images, as shown in Fig. 11. In addition, optimizing computational efficiency will be a key priority to facilitate real-time deployment in industrial settings.

Overall, this study demonstrates that deep-learning-based segmentation, when combined with fine-tuning and domain-specific adaptations, is a powerful tool for enhancing the accuracy and reliability of semiconductor imaging and metrology.

5 Conclusion

In this work, we proposed a deep-learning-based framework specifically designed for inner and outer segmentations in SEM/TEM images, addressing the unique challenges posed by low contrast, complex object structures, and multi-material compositions in semiconductor manufacturing. By leveraging a transformer-based architecture and incorporating advanced loss functions, our approach significantly improves segmentation accuracy and metrology precision.

The results demonstrate that our model achieves a high correlation with ground truth measurements, with an improved R^2 score for height and width estimation compared with baseline methods. The integration of recall evaluation further confirms that the model effectively balances the detection of all relevant structures while minimizing false positives. The ability to accurately segment both inner and outer objects contributes to a more reliable and automated analysis pipeline, reducing manual intervention in semiconductor research and manufacturing.

The proposed framework provides a robust solution for automating metrology processes in semiconductor imaging. The precise delineation of both internal voids and external structures ensures reliable measurements, which are crucial for quality control, defect analysis, and manufacturing optimization. By reducing reliance on manual annotations, this approach enhances throughput and consistency, enabling more scalable deployment in industrial settings. The use of deep learning also opens possibilities for adaptive models that can continuously learn from new data, further improving accuracy over time.

Compared with classical segmentation techniques, our deep learning model demonstrates superior robustness against low contrast variations, noise artifacts, and boundary ambiguities. Traditional methods often struggle with over-segmentation or under-segmentation, particularly in cases where objects share similar intensity levels with the background. Our proposed enhancements, such as contrast-aware attention and boundary-sensitive loss functions, mitigate these limitations by dynamically adapting to local image properties and preserving fine-grained object boundaries.

Despite its strong performance, our approach still faces certain challenges. The reliance on deep learning models means that a sufficiently large and well-annotated dataset is required to achieve optimal generalization. Although we implemented domain-specific fine-tuning to reduce data dependency, the model's performance in highly SEM/TEM imaging conditions may require additional adaptation. Furthermore, computational efficiency remains a consideration, particularly when deploying high-resolution models in real-time industrial workflows. Future optimizations, such as model pruning and knowledge distillation, could enhance inference speed without compromising segmentation accuracy.

To further improve segmentation performance and adaptability, future work will explore the integration of self-supervised learning techniques, which could reduce dependency on manually annotated datasets. In addition, expanding the model to incorporate multi-modal information—such as combining SEM/TEM images with other imaging modalities—could provide better feature representations, improving segmentation accuracy in challenging cases. Another promising direction is the development of an active learning framework, where user feedback is incorporated into model training, refining segmentation accuracy with minimal additional annotation efforts.

Overall, this study presents a significant step forward in the automated analysis of semiconductor structures, offering a scalable, adaptable, and precise solution for metrology and segmentation tasks. The gap between traditional image processing techniques and state-of-the-art deep learning, this work paves the way for more advanced, fully automated microscopy image analysis tools.

Disclosures

The authors declare that they have no relevant financial interests in the paper and no other potential conflicts of interest to disclose.

Code and Data Availability

The source code utilized in this study can be obtained upon request. For further details, please contact Johann Foucher at contact@pollen-metrology.com.

Acknowledgments

The framework of this work was funded by Auvergne-Rhône-Alpes. The authors' use of AI tools (ChatGPT, OpenAI) to assist with language refinement and grammar corrections. All scientific content, experimental design, and results were developed only by the authors.

References

1. Y. Yang et al., "Application of machine learning-based metrology for writer main pole CD measurement by CDSEM," *Proc. SPIE* **12053**, 120531R (2022).
2. I. W. Sanou et al., "Semi-automatic tools for nanoscale metrology and annotations for deep learning automation on electron microscopy images," *Proc. SPIE* **12749**, 127490D (2023).
3. I. W. Sanou et al., "Deep learning contour-based method for semi-automatic annotation of manufactured objects in electron microscopy images," *J. Electron. Imaging* **33**(3), 031204 (2024).
4. C. Thon et al., "Artificial intelligence in process engineering," *Adv. Intell. Syst.* **3**(6), 2000261 (2021).
5. L. Reimer, "Scanning electron microscopy: physics of image formation and microanalysis," *Meas. Sci. Technol.* **11**(12), 1826–1826 (2000).
6. A. Ferchichi et al., "Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: a systematic literature review," *Ecol. Inf.* **68**, 101552 (2022).
7. I. W. Sanou et al., "Deep learning aided tool for fast and accurate segmentation of multi-part semiconductor features," *Proc. SPIE* **12955**, 1295523 (2024).
8. I. W. Sanou et al., "Metrology and segmentation in SEM/TEM imaging: accelerating semiconductor analysis through advanced deep learning techniques," *Proc. SPIE* **13426**, 134262N (2025).
9. L. Reimer, *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*, Vol. 36, Springer (2013).
10. T. F. Cootes et al., "Active shape models-their training and application," *Comput. Vision Image Understanding* **61**(1), 38–59 (1995).
11. S. R. Gunn and M. S. Nixon, "A robust snake implementation; a dual active contour," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(1), 63–68 (1997).
12. T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.* **10**(2), 266–277 (2001).
13. L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Process.* **38**(1), 99–112 (1994).
14. A. Bieniek and A. Moga, "An efficient watershed algorithm based on connected components," *Pattern Recognit.* **33**(6), 907–916 (2000).
15. O. Juan and Y. Boykov, "Active graph cuts," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit. (CVPR'06)*, IEEE, Vol. **1**, pp. 1023–1029 (2006).
16. Y. Boykov and G. Funka-Lea, "Graph cuts and efficient ND image segmentation," *Int. J. Comput. Vision* **70**(2), 109–131 (2006).
17. A. Azeroual and K. Afdel, "Fast image edge detection based on Faber Schauder wavelet and Otsu threshold," *Heliyon* **3**(12), e00485 (2017).
18. T. Ö. Onur, "Improved image denoising using wavelet edge detection based on Otsu's thresholding," *Acta Polytech. Hungar.* **19**(2), 79–92 (2022).
19. K. He et al., "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2961–2969 (2017).
20. L. Castrejon et al., "Annotating object instances with a polygon-RNN," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 5230–5238 (2017).

21. S. Peng et al., “Deep snake for real-time instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 8533–8542 (2020).
22. B. Cheng et al., “Masked-attention mask transformer for universal image segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 1290–1299 (2022).
23. J. Jain et al., “OneFormer: one transformer to rule universal image segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 2989–2998 (2023).
24. A. Kirillov et al., “Segment anything,” in *Proc. IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, pp. 4015–4026 (2023).
25. X. Zhao et al., “Fast segment anything,” arXiv:2306.12156 (2023).
26. J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, “A comprehensive review of yolo architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS,” *Mach. Learn. Knowl. Extract.* **5**(4), 1680–1716 (2023).
27. X. Wang et al., “Cut and learn for unsupervised object detection and instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 3124–3134 (2023).
28. M. Caron et al., “Emerging properties in self-supervised Vision Transformers,” in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, pp. 9650–9660 (2021).
29. O. Siméoni et al., “Localizing objects with self-supervised transformers and no labels,” arXiv:2109.14279 (2021).
30. M. Pijarowski et al., “Utilizing grounded SAM for self-supervised frugal camouflaged human detection,” *Proc. SPIE* **13039**, 1303909 (2024).
31. I. Sanou, D. Conte, and H. Cardot, “An extensible deep architecture for action recognition problem,” in *14th Int. Joint Conf. Comput. Vision, Imaging and Comput. Graphics Theory and Appl. (VISAPP 2019)* (2019).
32. R. Aversa et al., “The first annotated set of scanning electron microscopy images for nanoscience,” *Sci. Data* **5**, 180172 (2018).
33. Y. Yu et al., “MCA: multidimensional collaborative attention in deep convolutional neural networks for image recognition,” *Eng. Appl. Artif. Intell.* **126**, 107079 (2023).
34. A. Yang et al., “Multi-feature self-attention super-resolution network,” *Vis. Comput.* **40**(5), 3473–3486 (2024).

Isaac Wilfried Sanou received his PhD in automation, robotics, and signal processing from the University of Toulon, France, in 2022. He is currently a postdoctoral researcher at Clermont Auvergne University and a data scientist at Pollen Metrology. His research focuses on deep learning and image processing methods for electron microscopy image segmentation in semiconductor metrology.

Julien Baderot is the research team leader at Pollen Metrology, a leading company in smart process control software for the semiconductor industry. He received his PhD in sciences from the University of Grenoble Alpes, specializing in signal and image processing, in collaboration with GIPSA Lab. His effort established the foundation for semiconductor metrology frameworks. He holds an engineering degree in electronics and optics from Polytech Orleans.

Vincent Barra is a full professor of computer science at Clermont Auvergne University with nearly 20 years of academic experience. His research spans data analysis from methodological and applied perspectives, including image and video processing, mesh processing, computational geometry, and machine/deep learning. He has authored or co-authored over 50 peer-reviewed journal articles, 5 books, and more than 100 papers in international conferences.

Ali Hallal is a materials scientist and a team leader for application and support at Pollen Metrology (Grenoble, France). He partners with semiconductor clients on process control and defect detection. His recent effort applies AI and deep learning to metrology, improving imaging data analysis in manufacturing. Trained in spintronics, magnetism, and 2D materials, with a PhD in nanotechnology, he bridges research and deployment and has co-authored several publications.

Léo Mazauric holds an engineering degree in mathematics and data science from Polytech Clermont-Ferrand, France, in 2022. He was a data scientist at Pollen Metrology, where he focused on optimizing epitaxy processes using machine learning techniques. His effort involved developing predictive models to improve semiconductor manufacturing efficiency. Passionate about data-driven innovation, he aims to bridge advanced analytics, AI, and industrial applications, particularly in high-precision environments such as microelectronics.

Johann Foucher is the CEO and co-founder (2014) of Pollen Metrology, a leader in AI-driven smart process control software for high-performance materials. Previously, he was an assignee at IBM (NY) managing advanced process control projects and a researcher/project manager at CEA-Leti in semiconductor nanometrology. He has 70+ publications, numerous invited talks, and 9 patents. He holds a PhD (Université Grenoble Alpes) in plasma physics for CMOS gate etching and an engineering degree from Polytech Orléans.