

Deep learning contour-based method for semi-automatic annotation of manufactured objects in electron microscopy images

Isaac Wilfried Sanou,^{a,b,*} Julien Baderot,^b Stéphanie Bricq,^a Yannick Benzech,^a Franck Marzani,^a Sergio Martinez,^b and Johann Foucher^b

^aUniversité de Bourgogne, ImViA EA 7535 Laboratory, Dijon, France

^bPollen Metrology, Moirans, France

ABSTRACT. Precision characterization is fundamental to achieve expected performance in semiconductors where Moore's law pushes the boundaries to miniaturize components. To measure these attributes, deep learning models are used, which require manual annotation of several objects captured via electron microscopy. However, this annotation can be laborious and time-consuming. We propose a semi-automated method for annotating items in electron microscopy images, in an effort to be innovative, efficient, and reliable. Our approach involves identifying objects, enhancing boundaries with use of a unique loss function incorporating physical aspects from electron microscopy images. It greatly reduces the need for users to undertake the annotation model's training process. It also minimizes post-inference processing by delivering a ready-to-use model. The constrained dynamic match loss (C-DML) incorporates dynamic matching with horizontal/vertical symmetry constraints to address the distinct challenges presented by manufactured objects acquired by microscopy imaging. Metrology metrics from the contour predictions obtained with C-DML obtain a mean relative error (MRE) of <10% and a correlation coefficient surpassing 90% when compared with ground truth corresponding to manual annotations. Our experimental results demonstrate the superior performance of C-DML over both classical DML and state-of-the-art deep annotation models. An extensive investigation demonstrates the effectiveness of our approach on heterogeneous datasets, including diverse objects of different materials and shapes, leading to state-of-the-art measurement results. Additionally, we show with the experiments that we can obtain better performances with better hyperparameters and data augmentation. Furthermore, this investigation presents a technique for annotating electron microscopy images efficiently and sheds light on the essential elements that dictate the approach's overall efficacy.

© 2024 SPIE and IS&T [DOI: [10.1117/1.JEI.33.3.031204](https://doi.org/10.1117/1.JEI.33.3.031204)]

Keywords: annotation; metrology; electron microscopy images; deep learning; segmentation

Paper 231102SS received Oct. 5, 2023; revised Jan. 12, 2024; accepted Jan. 19, 2024; published Feb. 8, 2024.

1 Introduction

In research and development, engineers aim to enhance industrial processes to produce objects with precise dimensions and uniformity. This is especially important in nanoscale metrology,

*Address all correspondence to Isaac Wilfried Sanou, isaac.sanou@pollen-metrology.com, isaac-Wilfried.Sanou@u-bourgogne.fr

where the quality and quantity of measurements are crucial, and obtaining images can be costly. Electron microscopy is a tool used to control the manufacturing process and final product to ensure compliance with specifications. This study analyzes three approaches to object metrology, with manual measurements serving as the reference method. While manual measurements offer precision tailored to user needs, they are time-consuming. On the other hand, automatic algorithms provide swift analysis but are less adaptable to process variations and require specific development. The use of deep learning capabilities is considered a game-changer due to its ability to deliver rapid and accurate results. However, the costs associated with annotations remain higher than the development time of automatic algorithms, as illustrated in Fig. 1. This paper explores these considerations, highlighting the trade-offs and advantages involved in the pursuit of efficient object metrology in the nanoscale domain.

Annotations, as outlined in Fig. 2, are typically more labor-intensive than standalone measurements. This often leads users to seek optimization approaches to reduce time and ensure high quality in the annotations.

To tackle this challenge, we introduce a segmentation-assisted approach that employs a deep learning contour-based model. In metrology, accurate segmentation is crucial for precise measurements. The model is first trained on electron microscopy images. During inference, users are required to draw a bounding box around the desired objects. By using the detected object contours, users can proceed seamlessly to the measurement stage, as illustrated in Fig. 3. The aim is to create a versatile annotation tool that enables users to perform metrology.

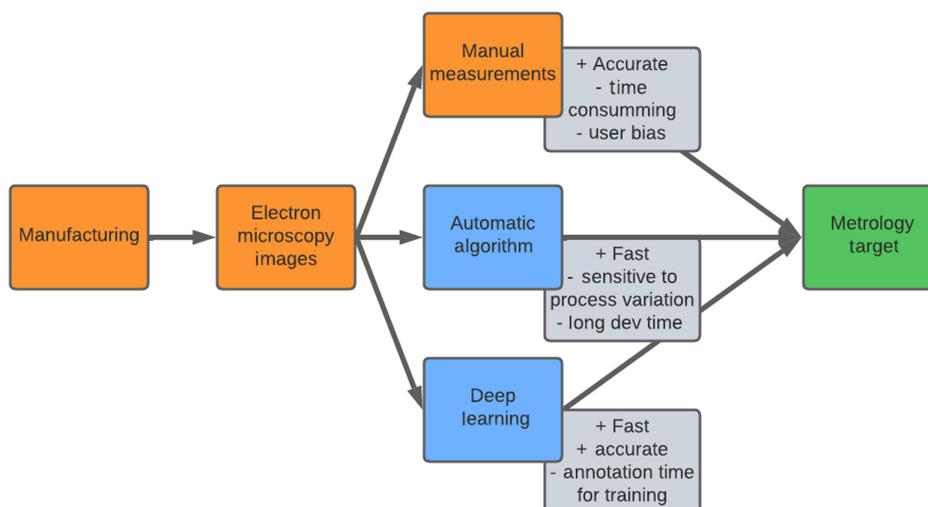


Fig. 1 Pipeline of metrology during research and development phase.

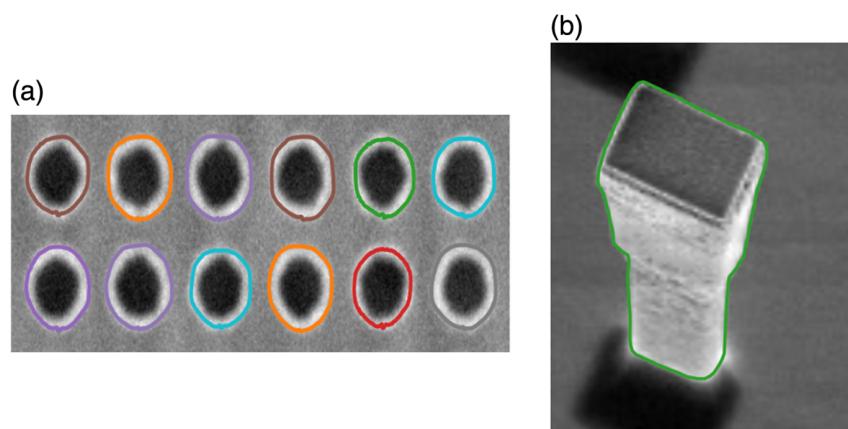


Fig. 2 Example of microscopy images annotation. (a) Image of contacts and (b) nanopillar.

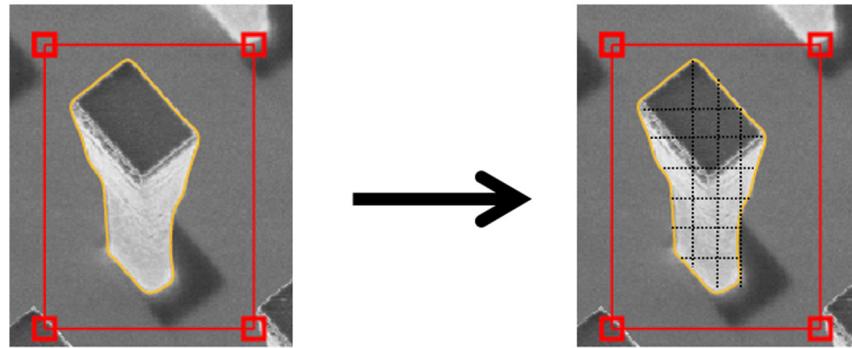


Fig. 3 Pipeline of annotation and metrology. Example on nanopillar object. Users draw first bounding box and after inference users can obtain measurements. Dot lines correspond to measurements ([Video 1, MP4, 6.60 MB \[URL: <https://doi.org/10.1117/1.JEI.33.3.031204.s1>\]](https://doi.org/10.1117/1.JEI.33.3.031204.s1)).

Later on, deep learning is used to extract segmentation, and then measurements are obtained by computing them from the strategy defined by the user. In this paper, we present maximum height and relative widths, which are two typical measurements used in object characterization. Our proposed deep learning pipeline simplifies user input and provides precise measurements by utilizing a strategic extraction approach. This approach ensures high-quality measurements by making segmentation the only criterion.

Various techniques to help the annotation of objects in images exist, including bounding boxes,^{1,2} polylines,³ modified segments,⁴ points,⁴ superpixel hierarchical annotations,⁵ polygons,⁶ interactive polygons,⁷ and keypoint annotations. Despite extensive exploration, most datasets for object instance localization rely on manual box or polygon annotations. State-of-the-art supervised instance segmentation methods, such as those employing Mask R-CNN,⁸ typically perform per-pixel segmentation within object bounding boxes detected.¹ In the same way, semantic segmentation⁹ is also used for pixel-wise annotation and it is usually preferred in an environmental context where the surroundings play an integral role such as Segnet¹⁰ and Unet3+.¹¹ In the literature, annotating methods can be broadly classified into three approaches.¹²

1.1 Classical Approaches

These techniques often involve energy minimization algorithms, like active contours,¹³ which deform an initial contour along object boundaries by energy optimization. Other variations, for instance, snakes combined with local shape models,¹⁴ and data-driven energy functions,¹⁵ have been developed. Watershed algorithms^{16,17} and graph cuts^{5,18} are other notable methods. Edge detection based on Otsu's thresholding is often used for annotation.^{19,20} These approaches are faster to implement but very sensitive to initialization.

1.2 Deep Learning Approaches

These methods leverage convolutional neural networks (CNNs). Common models include Mask R-CNN,⁸ PANet,²¹ CenterMask,²² YOLACT,²³ YOLACTedge,²⁴ SOLO,²⁵ Object Contour Detection,²⁶ Polygon RNN++,⁷ and Deep Snake.³ More recently, Meta researchers introduced segment anything model (SAM)²⁷ which aims to segment many types of images. Also, FastSAM²⁸ uses SAM training images with the corresponding masks and train on YOLO8 model. These approaches are typically fully supervised. Then, unsupervised models have been also developed such as Cut and Learn,²⁹ DINO,³⁰ and LOST³¹ which unfortunately have displayed lower performances with respect to their supervised counter parts. Both deep approaches take time for training but are more robust during inference.

1.3 Hybrid Approaches

These methods combine deep learning and classical methods. The models extract the first features using CNN and the use of a classical approach like snake algorithm is used to find the contour. We can mention deep active contour³² and learning deep structured active contours end-to-end,⁶ deep grabcut for object selection.³³ Most of these approaches have been benchmarked and outperformed a few deep learning approaches for annotation purposes. Deep

Snake architecture has been compared with Mask R-CNN, Polygon RNN++, and SGN³⁴ which gives state-of-the-art results.³ Recently, Deep Snake has also been compared in Soit,³⁵ Solov2,³⁶ and E2EC.³⁷

Contour detection in computer vision has shown the potential of deep learning over classical methods. Although classical methods are effective in many contexts, they may not be the best solution for microscopy images due to their unique characteristics, such as low-contrast edges and artifacts. Tailored solutions are necessary for optimal results as using deep model. Leveraging pre-trained weights from daily life datasets such as the KITTI INStance dataset,³⁸ the Semantic Boundaries Dataset,³⁹ the Cityscapes,⁴⁰ and the COCO dataset⁴¹ helps in selecting a generalized model. These architectures trained on daily life datasets are not suitable for microscopy images. The inherent challenge is due to the low contrast edges¹² and the image acquisition artifacts. Among the previous cited architectures, E2EC is selected because it was the most promising one on microscopy images based on empirical testing. SAM provides similar performances as E2EC model but it requires higher computing power to be fine-tuned on electron microscopy images.

In semiconductor applications, user-defined segmentation proves effective, eliminating the need to segment every element. Our research prioritizes practicality by providing a versatile annotation tool that minimizes user input, making it conducive for tasks like metrology and model training. The proposed approach seamlessly integrates deep learning, user-defined strategies, and prior shape knowledge to achieve precise measurements in microscopy images. This paper contributes valuable insights and a robust methodology for contour annotation in specialized domains.

The key contributions of this research paper are as follows:

- E2EC is fine-tuned on our annotation electron microscopy dataset.
- We have introduced a new loss function called Constrained Dynamic Match Loss (C-DML) to account for the physical priors of manufactured objects imaged by electron microscopy. This loss function incorporates a horizontal/vertical symmetry constraint, which we refer to as DML from E2EC.
- We benchmark our approach with state-of-the-art.
- Experiments on three different types of objects are conducted to choose the best set hyperparameters.
- We evaluate the accuracy of the segmentation with measurements for metrology with respect to the baseline which is the manual measurements.

The next sections of this paper are organized as follows: first, in Sec. 2, we explain the principle of the E2EC model. We propose our new model which takes the physical priors of manufactured objects on electron microscopy images. Then, in Sec. 3, we describe our dataset and the metrics used to evaluate the performance of the pipeline and the accuracy of the resulting measurements. We also show results and discuss the approach. Finally, Sec. 4 provides a conclusion and perspectives.

2 Methods: Semi-Automatic Microscopy Images Annotations

Our approach is based on E2EC architecture.³⁷ Even with only the COCO pre-trained model and no specific training on our dataset, the model gives promising results as visualized in Fig. 4.

The E2EC model presents an end-to-end contour-based instance segmentation approach that operates across multiple stages with high efficiency. It enables the use of a trainable contour architecture through the implementation of global deformation. In this context, global deformation involves regressing the contour vertices using only the center point features. Additionally, the architecture integrates multi-direction alignment to rectify the issue of divergence between the initialization and the anticipated-label vertex pairing. Dynamic matching loss (DML) is the main loss of E2EC as outlined in

$$\text{DML} = \frac{L_1(\text{pred}, \text{GT}) + L_2(\text{pred}, \text{GT})}{2}, \quad (1)$$

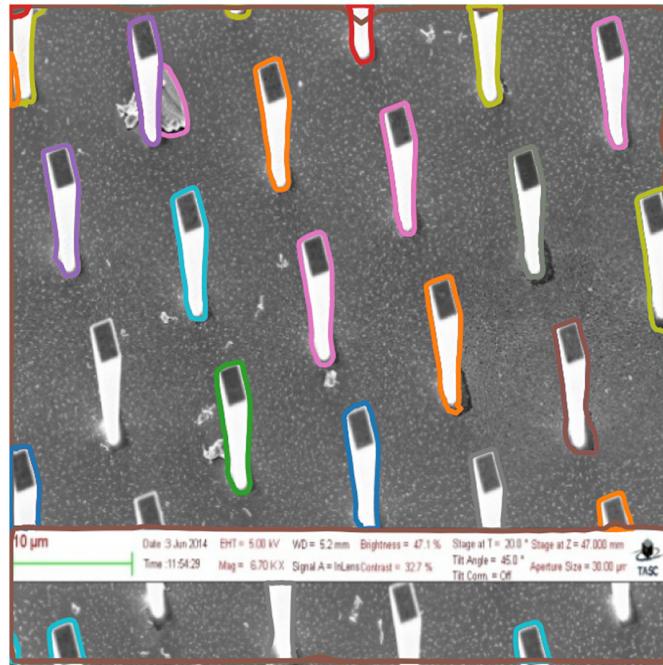


Fig. 4 Example of inference of E2EC on an electron microscopy image using pre-trained COCO dataset.

where L_1 , L_2 are the $\|\cdot\|_1$, $\|\cdot\|_2$, respectively, pred is the prediction, and GT is the ground truth. DML loss is started at epoch 10.

This loss aims to forecast the vertex points towards the nearest points on the label boundary. Moreover, it facilitates predicting the key label vertex while pulling the nearest predicted vertex towards its position. After a preliminary assessment on electron microscopy images, the architecture shows limitations such as the sensitivity to artifacts, linked to the geometric nature of some image databases used in our experiments as shown in Fig. 4. The electron microscopy data might include noisy or low-contrast images, factors for which E2EC is not robust. Consequently, we propose a novel approach by adapting E2EC. Notably, in the domain of metrology and electron microscopy images, the manufactured objects have certain shape characteristics that can be used as prior knowledge. For instance, object repetition and horizontal/vertical symmetry pushed us to impose a constraint on final loss, enhancing convergence and fit.

To impose a horizontal/vertical symmetry constraint on the resulting segmentation, we first define a symmetry metric. Consider an image A , to compute its horizontal/vertical symmetry [$\text{Sym}_v(A)$], we measure the mean square error (MSE) between A and its horizontally/vertically flipped version

$$\begin{aligned} \text{Sym}_h(A) &= \text{MSE}(A, \text{flip}_h(A)), \\ \text{Sym}_v(A) &= \text{MSE}(A, \text{flip}_v(A)). \end{aligned} \quad (2)$$

To incorporate this constraint, let GT denotes the ground truth and pred represents the model's prediction during training. The horizontal/vertical symmetry constraint takes the form as

$$\begin{aligned} \text{loss}_h(\text{GT}, \text{pred}) &= \|\text{Sym}_h(\text{GT}) - \text{Sym}_h(\text{pred})\|_2, \\ \text{loss}_v(\text{GT}, \text{pred}) &= \|\text{Sym}_v(\text{GT}) - \text{Sym}_v(\text{pred})\|_2. \end{aligned} \quad (3)$$

where $\text{Sym}(\text{GT})$ and $\text{Sym}(\text{pred})$ are given by Eq. (2). To consider an annotation tool adaptable to various microscopy image types, we encounter a complex dataset featuring some non-symmetric objects. To sidestep the need for dataset sorting prior to training because not all objects in our dataset are symmetric, we transform Eq. (3) into a proximal loss for horizontal/vertical symmetry constraint, expressed as

$$\text{proximal}^{\text{loss}_h} = \begin{cases} \text{loss}_h(\text{GT}, \text{pred}), & \text{if } \text{loss}_h < k_h, \text{ for horizontal symmetry constraint} \\ 0, & \text{else} \end{cases}, \quad (4)$$

$$\text{proximal}^{\text{loss}_v} = \begin{cases} \text{loss}_v(\text{GT}, \text{pred}), & \text{if } \text{loss}_v < k_v, \text{ for vertical symmetry constraint} \\ 0, & \text{else} \end{cases}, \quad (5)$$

where k_h and k_v are computed with all training set using Eq. (3). We then modify DML as shown in Eqs. (6) and (7) by adding symmetry constraint called constrained DML (C-DML)

$$C_{\text{DML}_h} = \text{DML} + \gamma_h * \text{proximal}^{\text{loss}_h}, \quad (6)$$

$$C_{\text{DML}_v} = \text{DML} + \gamma_v * \text{proximal}^{\text{loss}_v}, \quad (7)$$

where γ_v and γ_h are the penalty coefficient term. The penalty coefficient term permits to weight of the importance of the proximal loss. The more symmetrical the objects are, the more important it is to increase their value.

3 Experiments

This section presents our training and test datasets. Additionally, we provide metrology metrics, offering both quantitative and qualitative insights about performances of our approach. We show the results achieved and provide examples of inferences made on a public dataset.

3.1 Datasets and Metrics

Our training dataset is composed of both publicly available as the public SEM NFFA-Europe database⁴² and proprietary annotated images. Specifically, we annotated a total of 1501 electron microscopy images encompassing scanning electron microscopy (SEM) and transmission electron microscopy (TEM). SEM and TEM images are prevalent in research and development in the semiconductor area. SEM is commonly used for surface texture inspection of materials and rough material transitions. These images have in general low resolution and high noise. TEM offers insight into the internal structure of materials, including information about crystal structures and morphology. They are of better quality but are more expensive to produce and the preparation of the sample is destructive. The dataset is divided into two parts: one for training, consisting of 1244 images, and the other for testing, gathering 257 images. The test dataset is not used during training. It includes various types of pillars, wires, and gates. To assess the effectiveness of our approach, we apply metrology metrics illustrated in Fig. 5 which is more relevant for the industry than segmentation metrics. It is important to note that each image may contain multiple objects. For instance, pillars can contain more than five objects, wires can contain over 40 objects, and gates can also contain more than five objects. For each object type, they display varying designs, materials, microscopes, acquisition conditions, and aspects.

The performance of our approach uses two key metrics: the coefficient of correlation (r) and the mean relative error (MRE), defined in Eqs. (8) and (9), respectively

$$r = \frac{\sum_{i=1}^N (X_{\text{GT},i} - \bar{X}_{\text{GT}})(X_{\text{pred},i} - \bar{X}_{\text{pred}})}{\sqrt{\sum_{i=1}^N (X_{\text{GT},i} - \bar{X}_{\text{GT}})^2} \sqrt{\sum_{i=1}^N (X_{\text{pred},i} - \bar{X}_{\text{pred}})^2}}, \quad (8)$$

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \frac{|X_{\text{GT},i} - X_{\text{pred},i}|}{X_{\text{GT},i}}, \quad (9)$$

where $X_{\text{GT},i}$ and $X_{\text{pred},i}$ denote the ground truth and predicted measurements, respectively. N represents the number of dimensions. During the process, we consider four distinct measurements, including the maximal vertical object height (h_{\max}) and object width (w_x) measured at $x = 25\%$, $x = 50\%$, and $x = 75\%$ of the height from its bottom, as illustrated in Fig. 5. These measurements serve to determine the model performance across different aspects. The maximal height is obtained using internal geometric constraint system (GCS) tool, which allows for precise measurements based on predefined constraints on geometric objects. This tool facilitates the extraction of specific distances between two paths, in our case, the left and right segments.

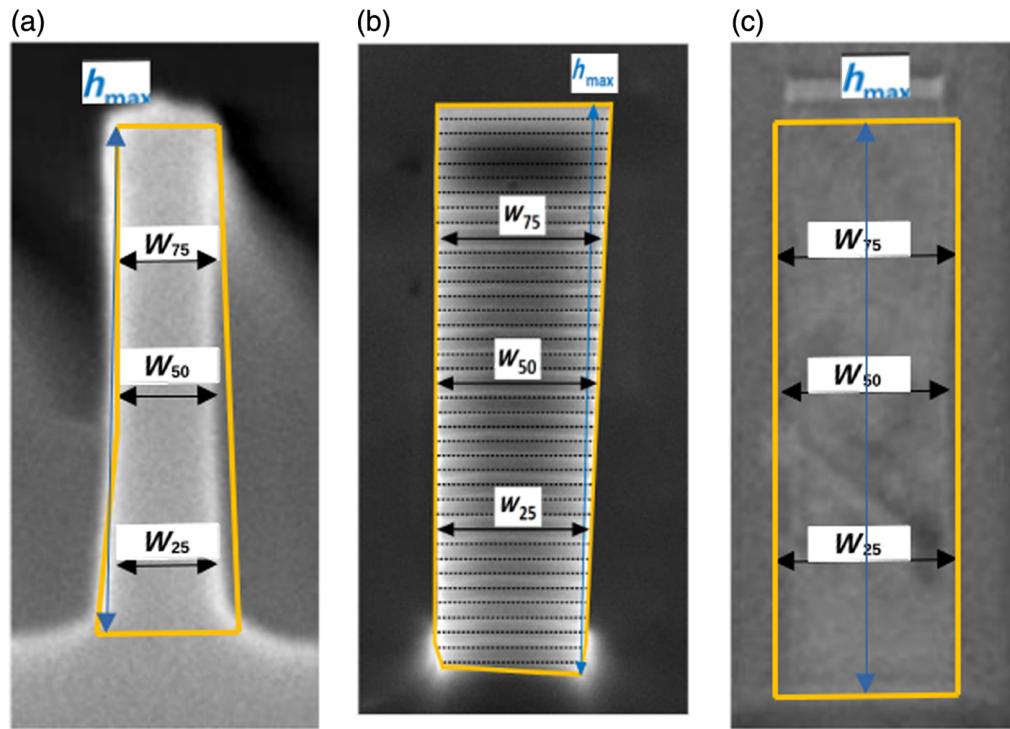


Fig. 5 Examples of images from (a) pillar, (b) wire, and (c) gate image. Illustration of the different measurements applied. h_{max} is the maximal vertical height. w_{25} , w_{50} , and w_{75} is the measure of the width of the object at 25%, 50%, and 75% of the height, respectively.

3.2 Results and Discussion

We present the initial outcomes of our contrasting technique with the modified E2EC and deep models approaches via electron microscopy imaging. Subsequent experiments enable us to determine more optimal hyperparameters for a standard semi-automatic annotation tool. The experimental phase is first to fine-tune the model using the complete 1244 images dataset using COCO pre-trained weights and adding our C-DML loss constrained. Second, we train the loss with different values of the coefficient penalty term (γ) to empirically determine the best coefficient. We start C-DML loss constrained at epoch 10 as in the E2EC paper.³⁷ In general, in semiconductor industry, a model or an algorithm could be considered reliable enough when achieving a coefficient correlation (r) exceeding 90% and an MRE below 10%.

3.2.1 Comparison of the proposed approach with deep contour-based models

Without retraining the models, we inferred E2EC and Deep Snake on our test set. E2EC and Deep Snake are deep contour-based models that can be extended as annotation tools. The aim of this study is to examine the performance of deep contour-based models on electron microscopy images. Table 1 presents the MRE and correlation of inference for E2EC and Deep Snake on the test set. The results indicate that E2EC performs well on images with pillars use cases, but fails on images with gates use cases.

Table 2 shows the MRE and correlation. Our proposed approach outperforms on the three test sets especially on pillars and wires use cases. In this experiment, we observed a correlation of 99% and an MRE of 2.9% for H_{max} for vertical C-DML. To achieve this, we used a $\gamma_v = 1$. Horizontal C-DML gives also better results than fine-tuning case and then E2EC with COCO pre-trained. Based on the shape of objects in our training set and the table, we determined that vertical symmetry is more suitable for our challenge than horizontal symmetry. We subsequently fine-tuned E2EC on our training dataset and used our proposed loss to further refine the model. To reach the best semi-automatic annotation, it is important to have a generic tool that provides accurate results across all test sets, not just certain ones by carefully selected hyperparameters.

Table 1 Comparison of E2EC and Deep Snake on an electron microscopy test set. Both models use a pre-trained model and are not retrained or fine-tuned. The Deep Snake algorithm was unsuccessful in providing clear outlines for the wires and gates cases.

Dataset		Pillars				Wires				Gates			
Method	Metrics	W_{25}	W_{50}	W_{75}	H_{\max}	W_{25}	W_{50}	W_{75}	H_{\max}	W_{25}	W_{50}	W_{75}	H_{\max}
E2EC	Corr Coeff (%)	95	94	88	98	54	67	89	90	19	28	30	80
Deep Snake	(%)	78	86	83	—	—	—	—	—	—	—	—	—
E2EC	MRE (%)	12.65	12.75	16.35	—	26.88	27.72	22.06	—	31.76	25.32	17.97	14.34
Deep Snake		31	17	21	—	—	—	—	—	—	—	—	—

Table 2 Comparison of E2EC, E2EC fine-tuning, E2EC fine-tuning incorporating horizontal C-DML and vertical C-DML on various electron microscopy datasets. $\gamma = 1$ in C-DML. Bold values indicate correlations above 90% and MRE below 10%.

Dataset		Pillars				Wires				Gates			
Method	Metrics	W_{25}	W_{50}	W_{75}	H_{\max}	W_{25}	W_{50}	W_{75}	H_{\max}	W_{25}	W_{50}	W_{75}	H_{\max}
E2EC	Corr Coeff (%)	95	95	88	95	54	67	89	95	19	28	30	80
E2EC Fine-tuning	(%)	98	98	98	99	98	98	98	98	98	98	98	98
Horizontal C-DML ($\gamma_h = 1$)		98	98	98	98	99	90	96	99	98	88	88	82
Vertical C-DML ($\gamma_v = 1$)		99	99	99	99	98	97	98	99	97	89	92	90
E2EC	MRE (%)	12.65	12.75	16.35	—	26.88	27.72	22.06	—	31.76	25.32	17.97	14.34
E2EC Fine-tuning		8.86	6.89	9.75	6.57	8.86	6.89	9.75	6.57	16.80	19.04	10.96	14.19
Horizontal C-DML ($\gamma_h = 1$)		8.73	9.54	15.07	6.4	10.62	10.22	9.67	5.37	14.58	16.84	12.38	13.71
Vertical C-DML ($\gamma_v = 1$)		7.93	7.10	8.90	2.9	7.93	7.10	8.90	2.9	11.62	12.10	10.38	13.71

3.2.2 Fine-tuning and choice of the best coefficient penalty term

The coefficient penalty term γ_v is an important parameter because it determines the significance of the constraint. Figure 6 shows a prediction of our approach and the COCO pre-trained weights. COCO pre-trained weights have double contour detected for certain objects and the contour does not follow the defined GT while, our approach detects one object and can handle the contour of the object.

Furthermore, in Fig. 7, an example of zeolite object inference is presented. The complexity of this use case is due to objects appearing in the background and overlapping with other objects. This inference demonstrates that our approach can be extended to other use cases with asymmetrical shapes.

Our approach shows strong performances for wire and pillar objects, displaying correlations above 90% and MRE below 10% between ground truth and prediction across diverse γ values, as depicted in Fig. 8. The yellow box plots is the baseline, it means that $\gamma = 0$ and this corresponds to make a fine-tuning of the model with the 1244 images of the dataset without symmetry

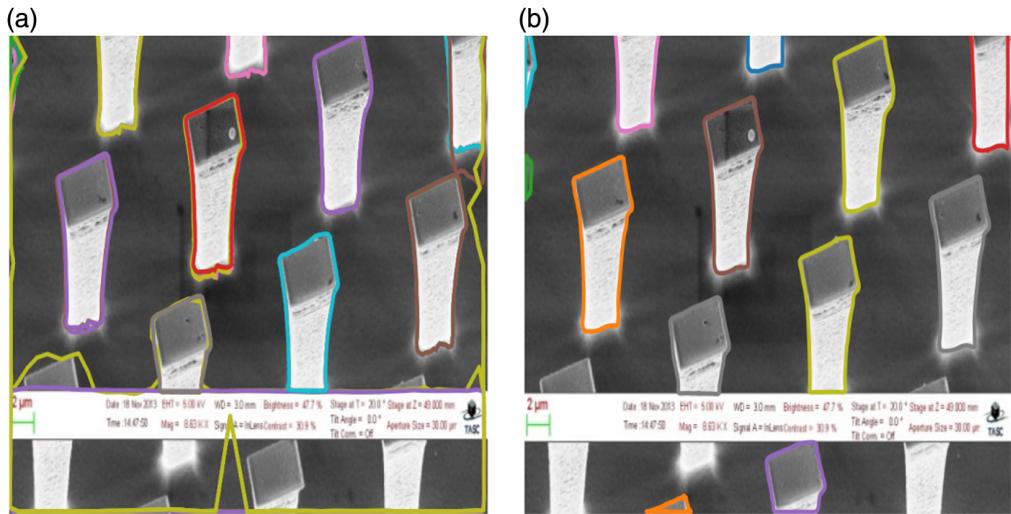


Fig. 6 Examples of the contours prediction for wire object. (a) COCO pre-trained inference. The Coco pre-trained model performs well in electron microscopy images but may require some additional post-processing. (b) Our approach inference. For our method, we do not need post-processing.

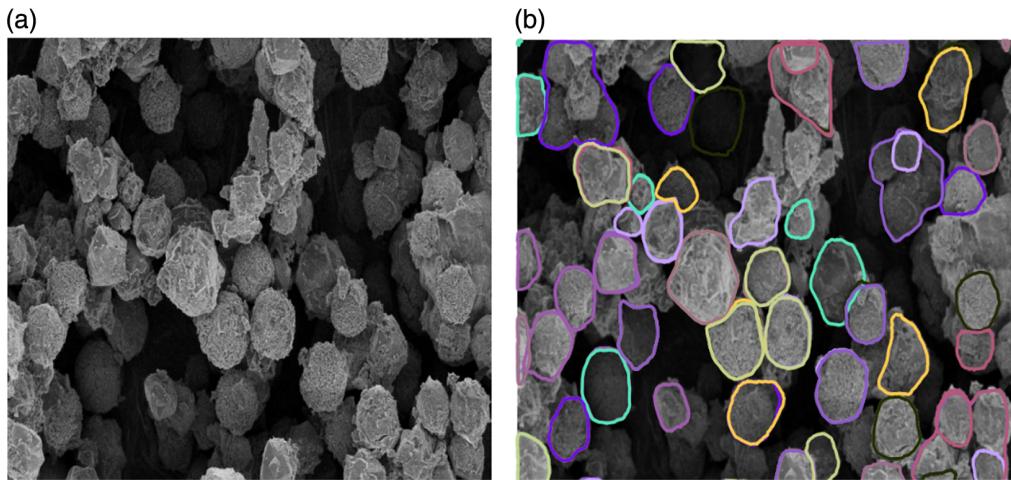


Fig. 7 Examples of the contours prediction for zeolites objects. (a) Original image and (b) our approach inference.

constraint. Results for gates are less encouraging for the industry, the metrics being outside the requirements.

In the case of pillars, r consistently falls within the 98% to 99% range for all γ values. Similarly, for wires, r spans between 95% and 99% across various γ values, excluding instances where $\gamma = 0.7$. These outcomes confirm the approach's capability to predict masks in proximity to the actual ground truth masks. However, for gate objects, the observed r values are insufficient, particularly at $\gamma = 0.7$.

Regarding MRE, both pillar and wire objects exhibit minimal measurement errors across various γ values. Notably, wires demonstrate superior performance at lower γ values, which can be attributed to the less pronounced symmetry in some wire objects as can be seen in Fig. 6.

Pillars dataset shows multiple symmetrical objects, highlighting acceptable outcomes achieved across diverse γ values. Conversely, in the case of gate dataset, the MRE becomes elevated as γ values increase, despite the symmetrical nature of gate objects. This challenge arises from the intrinsic complexities within the gates dataset. As shown in Fig. 5, gate objects display challenging contours to detect due to image acquisition factors. The limited visibility of contrast variations impacts contour detection, causing the model to struggle with correct contour

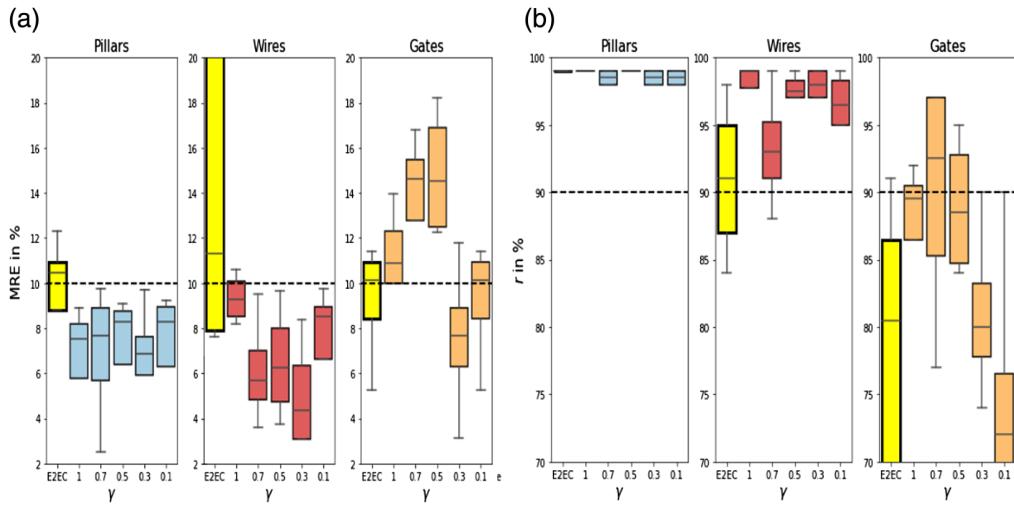


Fig. 8 (a) MRE for various values of γ in % for our C-DML. Dot line, the achievement value (10%). (b) Correlation coefficient for various values of γ in % for our C-DML. The baseline corresponding to E2EC fine-tuned on our train dataset is highlighted in yellow.

delineation. Consequently, it tends to overestimate maximum height when subjected to high vertical symmetric constraints. The metrics show the importance of the constraint applied to the loss function. For a generic model covering the three use cases, we identify a value of $\gamma = 0.5$.

3.2.3 Improvement of C-DML

This experiment is motivated by the fact that in the E2EC model, the starting epoch of DML loss function is fixed at 10. For that, we aim to find the best starting epoch of the loss.

Data augmentation is done during this experiment by rotating each image and the corresponding label to improve results.^{43,44} Let us consider as baseline, the previous result of MRE and r when $\gamma = 0.5$. In Fig. 9, we can compare the baseline, where the starting epoch is 10 without data augmentation against the result of the data augmentation. First, data augmentation shows the importance of having many data. The metrics MRE and r for each epoch of C-DML consistently outperform the baseline. Notably, there is an improvement in the results for gate objects. Furthermore, this experiment underscores the significance of the starting epoch for

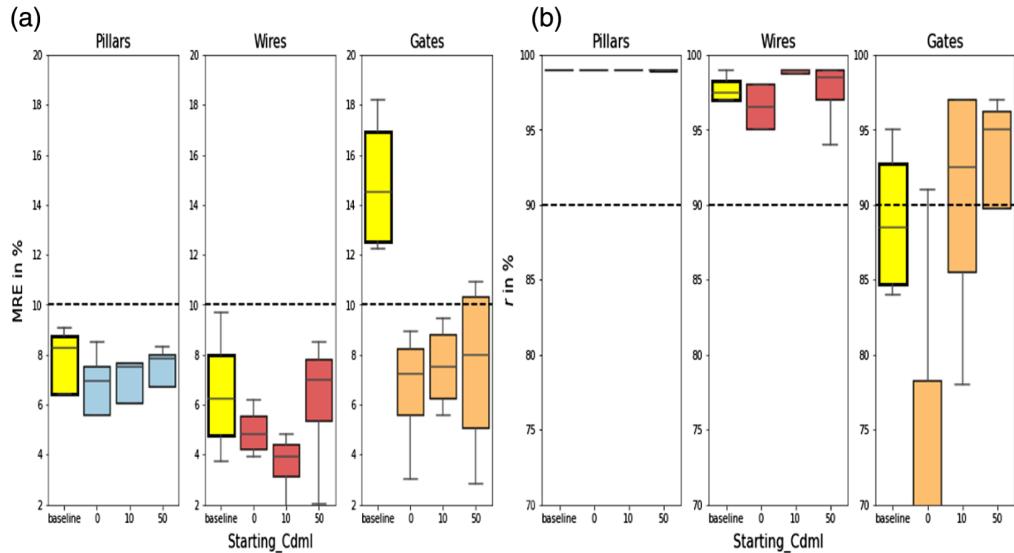


Fig. 9 (a) MRE of different epochs of starting C-DML in %. Dot line, the achievement value (10%). (b) Correlation coefficient of different epochs of starting C-DML in %. In yellow, the baseline with MRE and r when $\gamma = 0.5$ shown in Fig. 8.

C-DML. Notably, at epoch 0, the MRE and r for C-DML yield slightly inferior results due to its early initiation. We identify optimal performance with 10 epochs based on the empirical data.

4 Conclusion

We have presented a robust and efficient approach for annotating objects in electron microscopy images. This approach utilizes a deep learning-based contour model with a constrained loss function. Our method combines the strengths of existing state-of-the-art architectures, such as E2EC, and enhances them to accommodate the unique challenges inherent in electron microscopy images. The proposed approach not only identifies objects but also refines their boundaries using a novel loss function called C-DML. This loss function incorporates a vertical symmetry constraint that aligns with the physical characteristics of electron microscopy images.

Our experiments demonstrate the effectiveness of the proposed approach across three different types of datasets, including pillars, wires, and gates. We quantified the performance using coefficient correlation (r) and MRE metrics. The results indicate that the incorporation of the vertical symmetry constraint through C-DML improves the accuracy and robustness of the model. This is particularly evident in the case of symmetrical objects such as pillars, where the model achieves high correlations and low measurement errors.

Furthermore, we highlighted the importance of selecting appropriate hyperparameters and starting epochs for the C-DML loss during the training process. By carefully tuning these parameters, we were able to achieve better results compared to the baseline model.

In conclusion, our proposed approach provides accurate and reliable annotations, which are essential for metrology applications. Our work contributes to bridging the gap between deep learning-based techniques and the specific challenges posed by microscopy images, ultimately facilitating more efficient and accurate metrology analysis in semiconductor research and development. In future work, we aim to explore the application of our approach to other domains and datasets, further enhancing its adaptability and versatility. We also want to compare the training performances of supervised models using manual annotations versus the result of our tool.

Code and Data Availability

The data that support the findings of this study is available from the corresponding author upon reasonable request and with permission from Pollen Metrology.

Acknowledgments

We gratefully acknowledge the recovery plan funded by the National Research Agency under ANR-21-PRRD-0047-01.

References

1. R. Girshick, “Fast r-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1440–1448 (2015).
2. B. Adhikari et al., “Faster bounding box annotation for object detection in indoor scenes,” in *7th Eur. Workshop on Visual Inf. Process. (EUVIP)*, IEEE, pp. 1–6 (2018).
3. S. Peng et al., “Deep snake for real-time instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 8533–8542 (2020).
4. H. Li et al., “SIOD: single instance annotated per category per image for object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 14197–14206 (2022).
5. O. Juan and Y. Boykov, “Active graph cuts,” in *IEEE Comput. Soc. Conf. Comput. Vis. and Pattern Recognit. (CVPR’06)*, IEEE, Vol. 1, pp. 1023–1029 (2006).
6. D. Marcos et al., “Learning deep structured active contours end-to-end,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 8877–8885 (2018).
7. L. Castrejon et al., “Annotating object instances with a polygon-RNN,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 5230–5238 (2017).
8. K. He et al., “Mask r-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2961–2969 (2017).
9. B. Emek Soylu et al., “Deep-learning-based approaches for semantic segmentation of natural scene images: a review,” *Electronics* **12**(12), 2730 (2023).
10. V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: a deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).

11. H. Huang et al., “Unet 3+: a full-scale connected Unet for medical image segmentation,” in *ICASSP 2020-2020 IEEE Int. Conf. Acoust. Speech and Signal Process. (ICASSP)*, IEEE, pp. 1055–1059 (2020).
12. 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).
13. T. F. Cootes et al., “Active shape models-their training and application,” *Comput. Vis. Image Underst.* **61**(1), 38–59 (1995).
14. 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).
15. T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Trans. Image Process.* **10**(2), 266–277 (2001).
16. L. Najman and M. Schmitt, “Watershed of a continuous function,” *Signal Process.* **38**(1), 99–112 (1994).
17. A. Bieniek and A. Moga, “An efficient watershed algorithm based on connected components,” *Pattern Recognit.* **33**(6), 907–916 (2000).
18. Y. Boykov and G. Funka-Lea, “Graph cuts and efficient ND image segmentation,” *Int. J. Comput. Vis.* **70**(2), 109–131 (2006).
19. A. Azeroual and K. Afdel, “Fast image edge detection based on Faber Schauder wavelet and Otsu threshold,” *Heliyon* **3**(12), e00485 (2017).
20. T. Ö. Onur, “Improved image denoising using wavelet edge detection based on Otsu’s thresholding,” *Acta Polytech. Hung.* **19**(2), 79–92 (2022).
21. S. Liu et al., “Path aggregation network for instance segmentation,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 8759–8768 (2018).
22. Y. Lee and J. Park, “CenterMask: real-time anchor-free instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 13906–13915 (2020).
23. D. Bolya et al., “YOLACT: real-time instance segmentation,” in *Proc. IEEE/CVF Int. Conf. on Comput. Vis.*, pp. 9157–9166 (2019).
24. X. Wang et al., “SOLO: segmenting objects by locations,” *Lect. Notes Comput. Sci.* **12363**, 649–665 (2020).
25. H. Liu et al., “YolactEdge: real-time instance segmentation on the edge,” in *IEEE Int. Conf. Rob. and Autom. (ICRA)*, IEEE, pp. 9579–9585 (2021).
26. J. Yang et al., “Object contour detection with a fully convolutional encoder-decoder network,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 193–202 (2016).
27. A. Kirillov et al., “Segment anything,” arXiv:2304.02643 (2023).
28. X. Zhao et al., “Fast segment anything,” arXiv:2306.12156 (2023).
29. X. Wang et al., “Cut and learn for unsupervised object detection and instance segmentation,” arXiv:2301.11320 (2023).
30. M. Caron et al., “Emerging properties in self-supervised vision transformers,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 9650–9660 (2021).
31. O. Siméoni et al., “Localizing objects with self-supervised transformers and no labels,” arXiv:2109.14279 (2021).
32. C. Rupprecht et al., “Deep active contours,” arXiv:1607.05074 (2016).
33. N. Xu et al., “Deep grabcut for object selection,” arXiv:1707.00243 (2017).
34. S. Liu et al., “SGN: sequential grouping networks for instance segmentation,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3496–3504 (2017).
35. X. Yu et al., “SOIT: segmenting objects with instance-aware transformers,” in *Proc. AAAI Conf. Artif. Intell.*, Vol. 36, pp. 3188–3196 (2022).
36. X. Wang et al., “SOLOv2: dynamic and fast instance segmentation,” in *Adv. in Neural Inf. Process. Syst.*, Vol. 33, pp. 17721–17732 (2020).
37. T. Zhang, S. Wei, and S. Ji, “E2EC: an end-to-end contour-based method for high-quality high-speed instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 4443–4452 (2022).
38. L. Qi et al., “Amodal instance segmentation with KINS dataset,” in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 3014–3023 (2019).
39. M. Everingham et al., “The pascal visual object classes challenge: a retrospective,” *Int. J. Comput. Vis.* **111**(1), 98–136 (2015).
40. M. Cordts et al., “The cityscapes dataset for semantic urban scene understanding,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 3213–3223 (2016).
41. T.-Y. Lin et al., “Microsoft COCO: common objects in context,” *Lect. Notes Comput. Sci.* **8693**, 740–755 (2014).
42. R. Aversa et al., “The first annotated set of scanning electron microscopy images for nanoscience,” *Sci. Data* **5**(1), 1–10 (2018).
43. I. Sanou, D. Conte, and H. Cardot, “An extensible deep architecture for action recognition problem,” in *14th Int. Joint Conf. on Comput. Vis., Imaging and Comput. Graph. Theory and Appl. (VISAPP 2019)* (2019).

44. D. Tian et al., "Review of object instance segmentation based on deep learning," *J. Electron. Imaging* **31**(4), 041205 (2022).

Isaac Wilfried Sanou received his PhD in automation, robotics and signal processing in University of Toulon, France, in 2022. He was temporarily attached to education and research at University of Toulon – Engineering School, France. He is currently working as a postdoctoral researcher at University of Burgundy, France, and he is a data scientist at Pollen Metrology, France, since October 2022. His research interests include annotation, segmentation, machine learning, deep learning, and electron microscopy images in semiconductors.

Julien Baderot is the research tech leader at Pollen, which is a leading-edge company in smart process control software to accelerate the development of high-performance materials with its unique AI software technology. Before this position, he realized his PhD in sciences from the University Grenoble Alpes with a specialization in signal, image, parole, telecoms in collaboration with the GIPSA Lab to provide the foundation of a generic framework for the metrology in the semiconductor industry that is now a core technology in terms of image processing and machine learning. He holds an engineering degree in electronics and optics with a specialization in embedded vision from Polytech Orleans.

Stéphanie Bricq is associate professor at the ImViA Laboratory (University of Burgundy) since 2012. She defended her PhD in medical image processing in Strasbourg in 2008. Her main centers of interest are medical image processing and computer vision.

Yannick Benezeth is an associate professor in computer science at the University of Burgundy (France). He obtained his PhD in computer science from the University of Orléans in 2009. He also received his engineering degree from the ENSI de Bourges and his MS degree from the University of Versailles-Saint-Quentin-en-Yvelines in 2006. He also worked as a research fellow at the Orange Labs and INRIA (Rennes – France) between 2009 and 2011. His research interests include biomedical engineering, affective computing, image processing, and video analytics. Application areas include video health monitoring and endoscopy.

Franck Marzani obtained his PhD in computer vision and image processing in 1998. He has been a full professor at the University of Burgundy, Dijon, France, since 2009. His research interests include acquisition and analysis of images. He has been developing an activity on feature extraction from color and multispectral images for classification purposes. These methodologies have been proposed in the frame of different applications.

Sergio Martinez has joined Pollen as a data scientist in 2016. He has been actively involved in applying artificial intelligence for metrology in Smart3 product portfolio. After managing several research projects, director of research and application, Sergio is now VP of Technology. He holds a PhD in life and health sciences, focusing on using machine learning to detect early forms of cancer in the stomach.

Johann Foucher is the CEO of Pollen Metrology that he co-found in 2014. Pollen is a leading-edge company in smart process control software to accelerate the development of high performance materials with its unique AI software technology. Before Pollen, he was assignee at IBM (NY) as project manager in the field of advanced process control. Before that, he was a researcher and project manager at CEA/LETI institute in the field of nanometrology for semiconductor industry. He has published more than 70 papers, gave numerous invited talks at international conferences and has been granted nine patents. He holds a PhD in plasma physics applied to advanced semiconductor CMOS gate etching from University Grenoble Alpes and an engineering degree in electronics and plasmas physics from Polytech Orleans.